GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:11:32 ; Search time 4612.2 Seconds
(without alignments)
58.246 Million cell updates/sec

Title: US-09-396-196F-5
Perfect score: 25
Sequence: 1 gcgcagcaggtgcatcgccagcatt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	20.2	80.8	529	13	AG025369	AG025369 Oryza sat
c	2	19.8	79.2	506	11	BG266805	BG266805 1000101G1
	3	19.8	79.2	518	11	BI423547	BI423547 949049D11
	4	19.8	79.2	527	13	P977L	AL390686 Leishmani
c	5	19.8	79.2	582	10	AI665214	AI665214 605008C11
c	6	19.4	77.6	600	11	BF275487	BF275487 GA__Eb002
c	7	19.4	77.6	929	11	BG446019	BG446019 GA__Ea003
c	8	19.2	76.8	419	10	AI450725	AI450725 ms78c09.x
	9	19.2	76.8	495	10	AA170657	AA170657 ms78c09.r
	10	19.2	76.8	501	10	AI639758	AI639758 ms78c09.y
	11	19.2	76.8	562	13	AQ922633	AQ922633 RPC1-23-2
	12	19.2	76.8	671	11	BG605917	BG605917 RHIZ2_83_
	13	19.2	76.8	707	11	BF183133	BF183133 601809516
c	14	18.8	75.2	323	10	AU093350	AU093350 AU093350
c	15	18.8	75.2	385	10	AV436765	AV436765 AV436765
	16	18.8	75.2	475	10	AA821080	AA821080 GM09622.5
	17	18.8	75.2	516	10	AA264635	AA264635 LD08220.5
	18	18.8	75.2	541	13	TA80E01Q	AL461358 T. brucei
	19	18.8	75.2	637	13	AQ646542	AQ646542 RPCI93-Ec
	20	18.8	75.2	781	11	BG417267	BG417267 HVSMEk001
c	21	18.6	74.4	499	13	B76887	B76887 T27F2TF TAM
c	22	18.6	74.4	621	13	AQ969742	AQ969742 LERJP67TR
	23	18.6	74.4	659	13	AQ969741	AQ969741 LERJP67TF
	24	18.6	74.4	661	11	BI260142	BI260142 602968867
	25	18.6	74.4	708	10	BE513502	BE513502 601315005
	26	18.6	74.4	779	11	BE908400	BE908400 601503023
c	27	18.6	74.4	814	13	AQ576288	AQ576288 nbxb0088N
c	28	18.6	74.4	1101	13	CNS0020N	AL061904 Drosophil
	29	18.6	74.4	1332	11	BE871323	BE871323 601449431
c	30	18.4	73.6	343	10	AV397862	AV397862 AV397862
c	31	18.4	73.6	458	10	BE228836	BE228836 98AS3192
c	32	18.4	73.6	786	13	CNS03HMW	AL244481 Tetraodon
	33	18.2	72.8	165	10	AI901905	AI901905 618012B05
	34	18.2	72.8	290	10	AA741641	AA741641 LmLv39p3/
c	35	18.2	72.8	334	10	AU095285	AU095285 AU095285
c	36	18.2	72.8	349	11	BF753362	BF753362 QV0-CT058
	37	18.2	72.8	349	11	BG268511	BG268511 1000201E1
	38	18.2	72.8	355	10	BE397105	BE397105 601289536
	39	18.2	72.8	405	10	AW952110	AW952110 EST364180
	40	18.2	72.8	407	10	AI546214	AI546214 LD47729.5
c	41	18.2	72.8	420	10	AA387004	AA387004 vc20g03.r
	42	18.2	72.8	436	11	BF587149	BF587149 FM1_32_G0
	43	18.2	72.8	460	11	BF994016	BF994016 MR0-GN013
	44	18.2	72.8	470	10	BE268250	BE268250 601125693
	45	18.2	72.8	489	10	AI542824	AI542824 SD09292.5
	46	18.2	72.8	504	10	BE018236	BE018236 bb77d06.y
c	47	18.2	72.8	516	10	AJ284750	AJ284750 4A3B-AAB-
	48	18.2	72.8	527	13	AQ171736	AQ171736 HS_3172_B
	49	18.2	72.8	536	10	AI516136	AI516136 LD41893.5
	50	18.2	72.8	536	11	BF588135	BF588135 FM1_38_B0
	51	18.2	72.8	563	11	BG481153	BG481153 602528961
	52	18.2	72.8	563	11	BG257312	BG257312 602377834
	53	18.2	72.8	575	10	AA949558	AA949558 LD28409.5
	54	18.2	72.8	587	10	AI108331	AI108331 GH07146.5
	55	18.2	72.8	588	10	AI516564	AI516564 LD42465.5
	56	18.2	72.8	589	10	AA979043	AA979043 LD33291.5
	57	18.2	72.8	589	10	AI456290	AI456290 LD36182.5
	58	18.2	72.8	591	10	AA979205	AA979205 LD33523.5
	59	18.2	72.8	591	10	AI546543	AI546543 SD02890.5
	60	18.2	72.8	592	11	BI374927	BI374927 RE62620.5
	61	18.2	72.8	593	10	AI543819	AI543819 LD38951.5
	62	18.2	72.8	594	10	AA948864	AA948864 LD27534.5
	63	18.2	72.8	599	11	BI229714	BI229714 RE27550.5
	64	18.2	72.8	600	10	AI542478	AI542478 SD08817.5
	65	18.2	72.8	601	11	BI167238	BI167238 RE07574.5
	66	18.2	72.8	602	10	AI512930	AI512930 LD45062.5
	67	18.2	72.8	602	11	BI234055	BI234055 RE29685.5
	68	18.2	72.8	617	11	BI167745	BI167745 RE08202.5
	69	18.2	72.8	617	11	BI169400	BI169400 RE10564.5
	70	18.2	72.8	617	11	BI234713	BI234713 RE30461.5
	71	18.2	72.8	617	11	BI235805	BI235805 RE31738.5
	72	18.2	72.8	617	11	BI374345	BI374345 RE61908.5
	73	18.2	72.8	618	11	BI170181	BI170181 RE11559.5
	74	18.2	72.8	619	11	BI173817	BI173817 RE16931.5
	75	18.2	72.8	619	11	BI366322	BI366322 RE51833.5
	76	18.2	72.8	621	11	BI163314	BI163314 RE02679.5
	77	18.2	72.8	626	11	BI169865	BI169865 RE11162.5
	78	18.2	72.8	626	11	BI372477	BI372477 RE59449.5
	79	18.2	72.8	632	11	BI213668	BI213668 RE19118.5
	80	18.2	72.8	634	13	AQ350006	AQ350006 RPCI11-11
	81	18.2	72.8	636	10	AA942124	AA942124 LD26171.5
	82	18.2	72.8	637	11	BI232099	BI232099 RE28114.5
	83	18.2	72.8	638	11	BI173527	BI173527 RE16461.5
	84	18.2	72.8	639	11	BI240714	BI240714 RE37628.5
	85	18.2	72.8	654	10	AA950216	AA950216 LD30180.5

86	18.2	72.8	656	10	AA949604	AA949604 LD29562.5
87	18.2	72.8	663	11	BI228139	BI228139 RE25584.5
88	18.2	72.8	666	10	AA941547	AA941547 LD25735.5
89	18.2	72.8	670	11	BI228153	BI228153 RE25602.5
90	18.2	72.8	675	10	AA951507	AA951507 LD31994.5
91	18.2	72.8	677	10	AA817390	AA817390 LD22656.5
92	18.2	72.8	677	10	AA949251	AA949251 LD28075.5
93	18.2	72.8	678	11	BI353492	BI353492 GM23870.5
94	18.2	72.8	680	10	AA949098	AA949098 LD27860.5
95	18.2	72.8	685	10	AA941515	AA941515 LD25685.5
96	18.2	72.8	695	11	BI352111	BI352111 GM17686.5
97	18.2	72.8	695	11	BI374005	BI374005 RE61504.5
98	18.2	72.8	712	10	AV699372	AV699372 AV699372
99	18.2	72.8	723	10	AA820550	AA820550 LD24342.5
100	18.2	72.8	732	11	BG036589	BG036589 602326374

ALIGNMENTS

```
RESULT   1
AG025369/c
LOCUS       AG025369      529 bp    DNA             GSS       02-MAY-2000
DEFINITION  Oryza sativa DNA, 3' flanking sequence of Tos17 insertion,
            clone:T7115T, genomic survey sequence.
ACCESSION   AG025369
VERSION     AG025369.1  GI:7684033
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Oryza sativa (sub_species:japonica, strain:ND0016,
            cultivar:Nipponbare) DNA, clone_lib:PCR product directly amplified
            from rice genomic DNA clone:T7115T.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (sites)
  AUTHORS   Miyao,A. and Hirochika,H.
  TITLE     Rice insertion mutants
  JOURNAL   Unpublished (1999)
REFERENCE   2  (bases 1 to 529)
  AUTHORS   Miyao,A., Miyazaki,A., Yamashita,Y. and Hirochika,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-1999) to the DDBJ/EMBL/GenBank databases. Akio
            Miyao, National Institute of Agrobiological Resources, Molecular
            Genetics; 2-1-2, Kannondai, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:miyao@affrc.go.jp, URL:http://www.abr.affrc.go.jp/,
            Tel:81-298-38-7006, Fax:81-298-38-7006)
FEATURES             Location/Qualifiers
     source          1. .529
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /strain="ND0016"
                     /sub_species="japonica"
                     /db_xref="taxon:4530"
                     /clone="T7115T"
                     /clone_lib="PCR product directly amplified from rice
                     genomic DNA"
                     /note="Sequence group name: T7115T.-The 3' end of
                     retrotransposon Tos17 was found immediately upstream of
                     this sequence."
BASE COUNT      127 a    110 c    119 g    173 t
ORIGIN
```

```
  Query Match             80.8%;  Score 20.2;  DB 13;  Length 529;
  Best Local Similarity   88.0%;  Pred. No. 4e+02;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagcatt 25
           |||||||||  ||||  |||||||||
Db     239 GCGCAGCAGCAGCATTGCCAGCATT 215
```

```
RESULT   2
BG266805/c
LOCUS       BG266805      506 bp    mRNA            EST       20-FEB-2001
DEFINITION  1000101G12.x1 1000 - Unigene I from Maize Genome Project Zea mays
            cDNA, mRNA sequence.
ACCESSION   BG266805
VERSION     BG266805.1  GI:12970082
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 506)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 1000101  row: G  column: 12.
FEATURES             Location/Qualifiers
     source          1. .506
                     /organism="Zea mays"
                     /db_xref="dbEST:605008C11.x1"
                     /db_xref="taxon:4577"
                     /clone_lib="1000 - Unigene I from Maize Genome Project"
                     /note="This library represents the unique ESTs found in
                     the first round of EST sequencing at Stanford University
                     for the maize genome project. Sequences are present from
                     libraries 486, 487, 496, 603, 605, 606, 614, 618, 660, 683
                     , 687, 707, and 945. Contigs were assembled using TIGR's
                     CAP program and a representative EST from each contig was
                     selected for the Unigene set. All singlets were also
                     selected."
BASE COUNT       96 a    168 c    129 g    113 t
ORIGIN
```

```
  Query Match             79.2%;  Score 19.8;  DB 11;  Length 506;
  Best Local Similarity   91.3%;  Pred. No. 5.6e+02;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 gcagcaggtgcatcgccagcatt 25
           |||||||||||||| |||||| |
Db     459 GCAGCAGGTGCATCACCAGCAGT 437
```

```
RESULT   3
BI423547
LOCUS       BI423547      518 bp    mRNA            EST       16-AUG-2001
DEFINITION  949049D11.y1 949 - Juvenile leaf and shoot cDNA from Steve Moose
            Zea mays cDNA, mRNA sequence.
ACCESSION   BI423547
VERSION     BI423547.1  GI:15198896
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 518)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
```

Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949049 row: D column: 11.

FEATURES Location/Qualifiers
source 1. .518
/organism="Zea mays"
/cultivar="W64A"
/db_xref="taxon:4577"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve Moose"
/tissue_type="immature leaf primordium and vegetative meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XLOLR"
/note="Organ: juvenile vegetative shoots; Vector: pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts of total RNA by weight from 4 tissue sources (see below) were pooled, polyA+ RNA isolated, and cDNA synthesized for EcoRI (5') and XhoI (3') directional cloning into lambda Hybrizap vector from Stratagene. Tissue Sources: 1. Whole shoots 3 days after sowing/imbibing in wet soil. 2. Basal 1.5 cm shoots 6 days after sowing - includes yellow portions of developing leaves 1-5, primordia from 6-8, and the vegetative apex. 3. Non-green portions of developing leaves 4-5 and the vegetative apex, including adult leaf primordia, 9 days after sowing. 4. Partially expanded and greening leaves 4-5 at 13 days after sowing."

BASE COUNT 108 a 175 c 165 g 70 t
ORIGIN

Query Match 79.2%; Score 19.8; DB 11; Length 518;
Best Local Similarity 91.3%; Pred. No. 5.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      3 gcagcaggtgcatcgccagcatt 25
          ||||||||||||||| |||||| |
Db    236 GCAGCAGGTGCATCACCAGCAGT 258
```

RESULT 4
P977L
LOCUS P977L 527 bp DNA GSS 25-JUL-2000
DEFINITION Leishmania major Friedlin PAC P977 left end-sequence, similar to TR:Q9VH71 Q9VH71 CG12946 PROTEIN. [0] +2. . ., N=56, Prob=8.8e-14, genomic survey sequence.
ACCESSION AL390686
VERSION AL390686.1 GI:9501662
KEYWORDS GSS.
SOURCE Leishmania major.
ORGANISM Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
REFERENCE 1 (bases 1 to 527)
AUTHORS Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and Smith,D.F.
TITLE A physical map of the Leishmania major Friedlin genome
JOURNAL Genome Res. 8 (2), 135-145 (1998)
MEDLINE 98146435
REFERENCE 2 (bases 1 to 527)
AUTHORS Taylor,R.G., Huckle,E.E.J., Ivens,A.C., Rajandream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and alicat@sanger.ac.uk
COMMENT see http://www.ebi.ac.uk/parasites/leish.html
Details of Leishmania sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/L_major/
The primer sequence can be obtained from alicat@sanger.ac.uk.

FEATURES Location/Qualifiers
source 1. .527
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="PAC P977"

BASE COUNT 76 a 197 c 169 g 85 t
ORIGIN

Query Match 79.2%; Score 19.8; DB 13; Length 527;
Best Local Similarity 91.3%; Pred. No. 5.6e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 gcgcagcaggtgcatcgccagca 23
          || |||||||||| |||||||||
Db    277 GCTCAGCAGGTGCGTCGCCAGCA 299
```

RESULT 5
AI665214/c
LOCUS AI665214 582 bp mRNA EST 02-FEB-2000
DEFINITION 605008C11.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays cDNA, mRNA sequence.
ACCESSION AI665214
VERSION AI665214.2 GI:4804415
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 582)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT On May 11, 1999 this sequence version replaced gi:4776211.
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605008 row: C column: 11.
FEATURES Location/Qualifiers
source 1. .582
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5(alpha)"
/note="Organ: Kernel; Vector: pAD-GAL4-2'; Site_1: EcoRI; Site_2: XhoI; Kernel endosperm cDNA library from Schmidt lab"
BASE COUNT 115 a 191 c 146 g 130 t
ORIGIN

Query Match 79.2%; Score 19.8; DB 10; Length 582;
Best Local Similarity 91.3%; Pred. No. 5.7e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      3 gcagcaggtgcatcgccagcatt 25
          ||||||||||||||| |||||| |
Db    486 GCAGCAGGTGCATCACCAGCAGT 464
```

RESULT 6

```
BF275487/c
LOCUS       BF275487      600 bp    mRNA            EST       07-MAR-2001
DEFINITION  GA__Eb0024C21f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Eb0024C21f, mRNA sequence.
ACCESSION   BF275487
VERSION     BF275487.1  GI:11206557
KEYWORDS    EST.
SOURCE      Gossypium arboreum.
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
            ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence start: 3
            High quality sequence stop: 584.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Gossypium arboreum"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Eb0024C21f"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      170 a     94 c    160 g    176 t
ORIGIN


  Query Match             77.6%;  Score 19.4;  DB 11;  Length 600;
  Best Local Similarity   95.2%;  Pred. No. 8.1e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 gcagcaggtgcatcgccagca 23
           ||||||||||||||| ||||||
Db     282 GCAGCAGGTGCATCACCAGCA 262


RESULT   7
BG446019/c
LOCUS       BG446019      929 bp    mRNA            EST       15-MAR-2001
DEFINITION  GA__Ea0030G22f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Ea0030G22f, mRNA sequence.
ACCESSION   BG446019
VERSION     BG446019.1  GI:13355671
KEYWORDS    EST.
SOURCE      Gossypium arboreum.
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 929)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
            ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence start: 3
            High quality sequence stop: 428.
FEATURES             Location/Qualifiers
     source          1. .929
                     /organism="Gossypium arboreum"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Ea0030G22f"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      259 a     114 c     367 g     187 t       2 others
ORIGIN


  Query Match             77.6%;  Score 19.4;  DB 11;  Length 929;
  Best Local Similarity   95.2%;  Pred. No. 8.5e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 gcagcaggtgcatcgccagca 23
           ||||||||||||||| ||||||
Db     290 GCAGCAGGTGCATCACCAGCA 270


RESULT   8
AI450725/c
LOCUS       AI450725      419 bp    mRNA            EST       09-MAR-1999
DEFINITION  ms78c09.x1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:617680
            3', mRNA sequence.
ACCESSION   AI450725
VERSION     AI450725.1  GI:4299009
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            High quality sequence stop: 418.
FEATURES             Location/Qualifiers
     source          1. .419
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:617680"
                     /clone_lib="Soares mouse 3NbMS"
                     /sex="male"
```

/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTGTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 133 a 86 c 67 g 133 t
ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 419;
Best Local Similarity 87.5%; Pred. No. 9.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gcgcagcaggtgcatcgccagcat 24
| ||||||||||||| | ||||||
Db 297 GTGCAGCAGGTGCATAGACAGCAT 274

RESULT 9
AA170657
LOCUS AA170657 495 bp mRNA EST 16-FEB-1997
DEFINITION ms78c09.r1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:617680 5', mRNA sequence.
ACCESSION AA170657
VERSION AA170657.1 GI:1749200
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 495)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:378504
Seq primer: -28M13 rev2 from Amersham.
FEATURES Location/Qualifiers
source 1. .495
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:617680"
/clone_lib="Soares mouse 3NbMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTGTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 151 a 80 c 101 g 163 t
ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 495;
Best Local Similarity 87.5%; Pred. No. 9.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gcgcagcaggtgcatcgccagcat 24
| ||||||||||||| | ||||||
Db 213 GTGCAGCAGGTGCATAGACAGCAT 236

RESULT 10
AI639758
LOCUS AI639758 501 bp mRNA EST 15-MAR-2000
DEFINITION ms78c09.y1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:617680 5', mRNA sequence.
ACCESSION AI639758
VERSION AI639758.1 GI:4702867
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 501)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the correct orientation)
MGI:378504
Seq primer: -40RP from Gibco
High quality sequence stop: 477.
FEATURES Location/Qualifiers
source 1. .501
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:617680"
/clone_lib="Soares mouse 3NbMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTGTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

```
BASE COUNT      149 a     83 c    102 g    165 t      2 others
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 10;  Length 501;
  Best Local Similarity   87.5%;  Pred. No. 9.4e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcgcagcaggtgcatcgccagcat 24
          | ||||||||||||| | ||||||
Db    213 GTGCAGCAGGTGCATAGACAGCAT 236


RESULT  11
AQ922633
LOCUS       AQ922633      562 bp    DNA              GSS       21-DEC-1999
DEFINITION  RPCI-23-298P12.TJ RPCI-23 Mus musculus genomic clone RPCI-23-298P12
            , DNA sequence.
ACCESSION   AQ922633
VERSION     AQ922633.1  GI:6611731
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 562)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 298  row: P  column: 12
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .562
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-298P12"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).  "
BASE COUNT      169 a     92 c    121 g    180 t
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 13;  Length 562;
  Best Local Similarity   87.5%;  Pred. No. 9.5e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcgcagcaggtgcatcgccagcat 24
          | ||||||||||||| | ||||||
Db    486 GTGCAGCAGGTGCATAGACAGCAT 509


RESULT  12
BG605917
LOCUS       BG605917      671 bp    mRNA             EST       17-APR-2001
DEFINITION  RHIZ2_83_B10.g1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG605917
VERSION     BG605917.1  GI:13655888
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 671)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Seq primer: PolyTMix
            High quality sequence start: 44
            High quality sequence stop: 668
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .671
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      176 a    182 c    196 g    117 t
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 11;  Length 671;
  Best Local Similarity   87.5%;  Pred. No. 9.7e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 cgcagcaggtgcatcgccagcatt 25
          |||| |||||||||| |||||| |
Db    208 CGCAACAGGTGCATCACCAGCAGT 231


RESULT  13
BF183133
LOCUS       BF183133      707 bp    mRNA             EST       31-OCT-2000
DEFINITION  601809516F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4040306 5',
            mRNA sequence.
ACCESSION   BF183133
VERSION     BF183133.1  GI:11061320
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 707)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
```

```
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLCM857  row: p  column: 03
           High quality sequence stop: 613.
FEATURES             Location/Qualifiers
     source          1. .707
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4040306"
                     /clone_lib="NIH_MGC_18"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT      190 a    197 c    190 g    130 t
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 11;  Length 707;
  Best Local Similarity   87.5%;  Pred. No. 9.8e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcgcagcaggtgcatcgccagcat 24
          |||||||||| ||  |||||||||
Db    190 GCGCAGCAGGCGCCCCGCCAGCAT 213


RESULT  14
AU093350/c
LOCUS       AU093350      323 bp    mRNA            EST       21-JUN-2000
DEFINITION  AU093350 Rice callus Oryza sativa cDNA clone C0394, mRNA sequence.
ACCESSION   AU093350
VERSION     AU093350.1  GI:8594812
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 323)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from callus (2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
            C0394_8Z.
FEATURES             Location/Qualifiers
     source          1. .323
                     /organism="Oryza sativa"
                     /strain="cultivar Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="C0394"
                     /clone_lib="Rice callus"
                     /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
                     NotI; cDNA prepared from rice callus mRNAs by using
                     oligo(dT) as a primer and ligating to the SalI-NotI site
                     of pBluescript II SK+ phagemid.  "
BASE COUNT       59 a     97 c    112 g     53 t      2 others
ORIGIN


  Query Match             75.2%;  Score 18.8;  DB 10;  Length 323;
  Best Local Similarity   90.9%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 gcagcaggtgcatcgccagcat 24
          |||||||| ||||| |||||||
Db    266 GCAGCAGGCGCATCTCCAGCAT 245


RESULT  15
AV436765/c
LOCUS       AV436765      385 bp    mRNA            EST       23-AUG-2000
DEFINITION  AV436765 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PS010a09_r 5', mRNA sequence.
ACCESSION   AV436765
VERSION     AV436765.1  GI:8591990
KEYWORDS    EST.
SOURCE      Porphyra yezoensis.
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 385)
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .385
                     /organism="Porphyra yezoensis"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PS010a09_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       56 a     120 c    144 g     65 t
ORIGIN


  Query Match             75.2%;  Score 18.8;  DB 10;  Length 385;
  Best Local Similarity   90.9%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 gcgcagcaggtgcatcgccagc 22
          || |||||||||||| |||||||
Db    326 GCACAGCAGGTGCAGCGCCAGC 305


RESULT  16
AA821080
LOCUS       AA821080      475 bp    mRNA            EST       23-APR-2001
DEFINITION  GM09622.5prime GM Drosophila melanogaster ovary BlueScript
            Drosophila melanogaster cDNA clone GM09622 5prime, mRNA sequence.
ACCESSION   AA821080
VERSION     AA821080.1  GI:2890948
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 475)
```

```
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 96  row: B  column: 10
            High quality sequence stop: 439.
FEATURES             Location/Qualifiers
     source          1. .475
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GM09622"
                     /clone_lib="GM Drosophila melanogaster ovary BlueScript"
                     /sex="female"
                     /dev_stage="newly eclosed females: germarium-stage 6"
                     /lab_host="SOLR"
                     /note="Organ: ovary; Vector: BlueScript SK; Site_1: EcoRI;
                     Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      128 a    147 c    121 g     79 t
ORIGIN
```

```
  Query Match             75.2%;  Score 18.8;  DB 10;  Length 475;
  Best Local Similarity   90.9%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagc 22
           ||| || ||||||||||||||||
Db     337 GCGGAGGAGGTGCATCGCCAGC 358
```

```
RESULT  17
AA264635
LOCUS       AA264635      516 bp    mRNA            EST       19-APR-2001
DEFINITION  LD08220.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD08220 5prime, mRNA sequence.
ACCESSION   AA264635
VERSION     AA264635.1  GI:1900679
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 82  row: B  column: 8
            High quality sequence stop: 515.
FEATURES             Location/Qualifiers
     source          1. .516
                     /organism="Drosophila melanogaster"
                     /db_xref="BDGP_EST:BDcin007567"
                     /db_xref="taxon:7227"
                     /clone="LD08220"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1: EcoRI
                     ; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT      145 a    154 c    138 g     79 t
ORIGIN
```

```
  Query Match             75.2%;  Score 18.8;  DB 10;  Length 516;
  Best Local Similarity   90.9%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagc 22
           ||||||| ||| ||||||||||||
Db     356 GCGCAGGAGGAGCATCGCCAGC 377
```

```
RESULT  18
TA80E01Q
LOCUS       TA80E01Q      541 bp    DNA             GSS       13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 80e01, reverse sequence,
            genomic survey sequence.
ACCESSION   AL461358
VERSION     AL461358.1  GI:11860640
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .541
                     /organism="Trypanosoma brucei"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="80e01"
BASE COUNT      113 a    141 c    151 g    136 t
ORIGIN
```

```
  Query Match             75.2%;  Score 18.8;  DB 13;  Length 541;
  Best Local Similarity   90.9%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagc 22
           |||| ||||| ||||||||||||
Db     481 GCGCCGCAGGAGCATCGCCAGC 502
```

```
RESULT  19
AQ646542
LOCUS       AQ646542      637 bp    DNA             GSS       08-JUL-1999
```

```
DEFINITION  RPCI93-EcoRI-6M17.TJ RPCI93-EcoRI Trypanosoma brucei genomic clone
            RPCI93-EcoRI-6M17, DNA sequence.
ACCESSION   AQ646542
VERSION     AQ646542.1  GI:5123252
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 637)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.
            , Fraser,C. and Adams,M.
  TITLE     Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPCI-93
            Library for gene discovery and sequence-ready map construction
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI93-EcoRI-6M17.TV
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones and high density filters may be purchased from BACPAC
            Resources (http://bacpac.med.buffalo.edu. BAC end sequences search
            page: http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .637
                     /organism="Trypanosoma brucei"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="RPCI93-EcoRI-6M17"
                     /clone_lib="RPCI93-EcoRI"
                     /note="Vector: pBACe3.6; Site_1: Eco RI; Site_2: Eco RI;
                     Constructed for The Institute for Genomic Research by
                     Bohui Zhao in Pieter de Jong's laboratory (Roswell Park
                     Cancer Institute, Buffalo, NY). Briefly, Trypanosoma
                     brucei TREU927/4 GUTat 10.1 agarose embedded DNA was
                     partially digested with a combination of Eco RI and Eco RI
                     methylase (RPCI93-EcoRI segment) or Dpn II (RPCI93-DpnII
                     segment).  High molecular weight fragments were ligated in
                     pBACe3.6 vector digested with Eco RI or Bam H1,
                     respectively.  The average insert size is 141 Kb.  Total
                     coverage (both segments): > 90 X the haploid
                     non-minichromosomal genome."
BASE COUNT      141 a    157 c    189 g    150 t
ORIGIN
```

```
  Query Match             75.2%;  Score 18.8;  DB 13;  Length 637;
  Best Local Similarity   90.9%;  Pred. No. 1.4e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagc 22
           |||| ||||| ||||||||||
Db     132 GCGCCGCAGGAGCATCGCCAGC 153
```

```
RESULT  20
BG417267
LOCUS       BG417267       781 bp    mRNA            EST       13-MAR-2001
DEFINITION  HVSMEk0017A16f Hordeum vulgare testa/pericarp EST library
            HVcDNA0013 (normal) Hordeum vulgare cDNA clone HVSMEk0017A16f, mRNA
            sequence.
ACCESSION   BG417267
VERSION     BG417267.1  GI:13322818
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 781)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
            ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
            Wood,T.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 753.
FEATURES             Location/Qualifiers
     source          1. .781
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEk0017A16f"
                     /clone_lib="Hordeum vulgare testa/pericarp EST library
                     HVcDNA0013 (normal)"
                     /tissue_type="testa/pericarp"
                     /lab_host="TJC121"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; For
                     more details on library preparation and sequence analysis
                     see http://www.genome.clemson.edu/projects/barley/ To
                     order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT      191 a    211 c    253 g    126 t
ORIGIN
```

```
  Query Match             75.2%;  Score 18.8;  DB 11;  Length 781;
  Best Local Similarity   90.9%;  Pred. No. 1.4e+03;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 cagcaggtgcatcgccagcatt 25
           |||||||||||| |||||| |
Db     291 CAGCAGGTGCATCACCAGCAGT 312
```

```
RESULT  21
B76887/c
LOCUS       B76887         499 bp    DNA             GSS       16-JAN-1998
DEFINITION  T27F2TF TAMU Arabidopsis thaliana genomic clone T27F2, DNA
            sequence.
ACCESSION   B76887
VERSION     B76887.1  GI:2773526
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.
            , Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
            ,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
            Arabidopsis Genomic Sequencing. Update 3
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: T27F2TR
            Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
```

```
            Email: rounsley@tigr.org
            Seq primer: M13-21
            Class: BAC ends
            High quality sequence stop: 499.
FEATURES              Location/Qualifiers
     source           1. .499
                      /organism="Arabidopsis thaliana"
                      /strain="Columbia"
                      /db_xref="taxon:3702"
                      /clone="T27F2"
                      /clone_lib="TAMU"
                      /sex="hermaphrodite"
                      /note="Vector: BeloBACII; Site_1: HindIII; Site_2: HindIII
                      ; Produced by Rod Wing"
BASE COUNT       141 a      90 c     116 g     152 t
ORIGIN
```

```
  Query Match             74.4%;  Score 18.6;  DB 13;  Length 499;
  Best Local Similarity   84.0%;  Pred. No. 1.6e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagcatt 25
           | |||||||  |||| |||||||||
Db     218 GAGCAGCAGCGGCATTGCCAGCATT 194
```

```
RESULT  22
AQ969742/c
LOCUS       AQ969742      621 bp    DNA             GSS       28-JAN-2000
DEFINITION  LERJP67TR LERG Arabidopsis thaliana genomic clone LERJP67, DNA
            sequence.
ACCESSION   AQ969742
VERSION     AQ969742.1  GI:6797443
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Seq primer: TR
            Class: shotgun.
FEATURES              Location/Qualifiers
     source           1. .621
                      /organism="Arabidopsis thaliana"
                      /strain="Landsberg erecta"
                      /db_xref="taxon:3702"
                      /clone="LERJP67"
                      /clone_lib="LERG"
                      /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
                      sheared to 0.4-0.7 Kbp before ligation."
BASE COUNT       177 a     108 c     125 g     209 t       2 others
ORIGIN
```

```
  Query Match             74.4%;  Score 18.6;  DB 13;  Length 621;
  Best Local Similarity   84.0%;  Pred. No. 1.6e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy       1 gcgcagcaggtgcatcgccagcatt 25
           | |||||||  |||| |||||||||
Db     525 GAGCAGCAGCTGCATTTCCAGCATT 501
```

```
RESULT  23
AQ969741
LOCUS       AQ969741      659 bp    DNA             GSS       28-JAN-2000
DEFINITION  LERJP67TF LERG Arabidopsis thaliana genomic clone LERJP67, DNA
            sequence.
ACCESSION   AQ969741
VERSION     AQ969741.1  GI:6797442
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Seq primer: TF
            Class: shotgun.
FEATURES              Location/Qualifiers
     source           1. .659
                      /organism="Arabidopsis thaliana"
                      /strain="Landsberg erecta"
                      /db_xref="taxon:3702"
                      /clone="LERJP67"
                      /clone_lib="LERG"
                      /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
                      sheared to 0.4-0.7 Kbp before ligation."
BASE COUNT       222 a     136 c     114 g     187 t
ORIGIN
```

```
  Query Match             74.4%;  Score 18.6;  DB 13;  Length 659;
  Best Local Similarity   84.0%;  Pred. No. 1.6e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagcatt 25
           | |||||||  |||| |||||||||
Db     109 GAGCAGCAGCTGCATTTCCAGCATT 133
```

```
RESULT  24
BI260142
LOCUS       BI260142      661 bp    mRNA            EST       17-JUL-2001
DEFINITION  602968867F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5108345 5',
            mRNA sequence.
ACCESSION   BI260142
VERSION     BI260142.1  GI:14818167
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM11262  row: e  column: 18
           High quality sequence stop: 660.
FEATURES             Location/Qualifiers
     source          1. .661
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5108345"
                     /clone_lib="NIH_MGC_12"
                     /tissue_type="cervical carcinoma cell line"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.4 kb. Library prepared by Life
                     Technologies."
BASE COUNT      130 a    215 c    199 g    117 t
ORIGIN

  Query Match             74.4%;  Score 18.6;  DB 11;  Length 661;
  Best Local Similarity   84.0%;  Pred. No. 1.6e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagcatt 25
           ||||||||| | ||  |||||||||
Db     184 GCGCAGCAGCTCCAGGGCCAGCATT 208


RESULT  25
BE513502
LOCUS       BE513502      708 bp    mRNA            EST       07-AUG-2000
DEFINITION  601315005F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3633864 5',
            mRNA sequence.
ACCESSION   BE513502
VERSION     BE513502.1  GI:9720714
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 708)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM327  row: i  column: 01
            High quality sequence stop: 677.
FEATURES             Location/Qualifiers
     source          1. .708
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3633864"
                     /clone_lib="NIH_MGC_8"
                     /tissue_type="Burkitt lymphoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      138 a    192 c    267 g    111 t
ORIGIN

  Query Match             74.4%;  Score 18.6;  DB 10;  Length 708;
  Best Local Similarity   84.0%;  Pred. No. 1.7e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagcatt 25
           | | |||||||||||| ||||||| |
Db     309 GAGGAGCAGGTGCATGGCCAGCACT 333


RESULT  26
BE908400
LOCUS       BE908400      779 bp    mRNA            EST       20-OCT-2000
DEFINITION  601503023F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3905068 5',
            mRNA sequence.
ACCESSION   BE908400
VERSION     BE908400.1  GI:10402938
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 779)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9711  row: m  column: 05
            High quality sequence start: 4
            High quality sequence stop: 622.
FEATURES             Location/Qualifiers
     source          1. .779
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3905068"
                     /clone_lib="NIH_MGC_70"
                     /tissue_type="epithelioid carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.1 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      159 a    208 c    276 g    136 t
ORIGIN

  Query Match             74.4%;  Score 18.6;  DB 11;  Length 779;
  Best Local Similarity   84.0%;  Pred. No. 1.7e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagcatt 25
           | | |||||||||||| ||||||| |
Db     327 GAGGAGCAGGTGCATGGCCAGCACT 351


RESULT  27
AQ576288/c
LOCUS       AQ576288      814 bp    DNA             GSS       02-JUN-1999
```

```
DEFINITION  nbxb0088N14f CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0088N14f, DNA sequence.
ACCESSION   AQ576288
VERSION     AQ576288.1  GI:4976773
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 814)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence stop: 427.
FEATURES             Location/Qualifiers
     source          1. .814
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbxb0088N14f"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world  population especially those
                     inhabiting highly populated areas of the  humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a  haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991).  The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to  facilitate positional cloning, physical mapping
                     and genome sequencing  of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare  variety. The
                     library contains 36,864 clones with an average insert size
                     of 128.5 Kb providing 10.9 haploid genome equivalents. The
                     deep  coverage allows the isolation a particular sequence
                     with a probability  of 99.9 %. Two high density filters,
                     each containing 18,432 clones  (doubly spotted), represent
                     the whole library for colony screening."
BASE COUNT      190 a    166 c    202 g    249 t      7 others
ORIGIN
```

```
  Query Match             74.4%;  Score 18.6;  DB 13;  Length 814;
  Best Local Similarity   84.0%;  Pred. No. 1.7e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagcatt 25
           | ||||||| |||| ||||||||||
Db     418 GTGCAGCAGCAGCATTGCCAGCATT 394
```

```
RESULT  28
CNS0020N/c
LOCUS       CNS0020N     1101 bp    DNA              GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR05E21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL061904
VERSION     AL061904.1  GI:4940165
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR05E21"
                     /note="end : TET3"
BASE COUNT      275 a    202 c    223 g    319 t     82 others
ORIGIN
```

```
  Query Match             74.4%;  Score 18.6;  DB 13;  Length 1101;
  Best Local Similarity   84.0%;  Pred. No. 1.7e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagcatt 25
           ||| ||||| |||| ||||| ||||
Db     819 GCGTAGCAGCTGCACCGCCACCATT 795
```

```
RESULT  29
BE871323
LOCUS       BE871323     1332 bp    mRNA             EST       20-OCT-2000
DEFINITION  601449431F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3853614 5',
            mRNA sequence.
ACCESSION   BE871323
VERSION     BE871323.1  GI:10320099
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1332)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
```

```
           Plate: LLAM9577  row: m  column: 07
           High quality sequence stop: 381.
FEATURES             Location/Qualifiers
     source          1. .1332
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3853614"
                     /clone_lib="NIH_MGC_65"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies.  "
BASE COUNT      317 a    544 c    276 g    195 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 11;  Length 1332;
  Best Local Similarity   84.0%;  Pred. No. 1.8e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagcatt 25
           || |||||  ||| ||||||||||||
Db     352 GCCCAGCACATGCTTCGCCAGCATT 376


RESULT  30
AV397862/c
LOCUS       AV397862      343 bp    mRNA            EST       29-SEP-2000
DEFINITION  AV397862 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
            cDNA clone CM002c09_r, mRNA sequence.
ACCESSION   AV397862
VERSION     AV397862.1  GI:6552078
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 343)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
  TITLE     A large scale structural analysis of cDNAs in a unicellular green
            alga, Chlamydomonas reinhardtii. I. Generation of 3433
            non-redundant expressed sequence tags
  JOURNAL   DNA Res. 6 (6), 369-373 (1999)
  MEDLINE   20152988
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .343
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="CM002c09_r"
                     /clone_lib="Chlamydomonas reinhardtii C9"
                     /dev_stage="photoautotrophic growth"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       62 a    108 c    101 g     72 t
ORIGIN


  Query Match             73.6%;  Score 18.4;  DB 10;  Length 343;
  Best Local Similarity   95.0%;  Pred. No. 1.8e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgcca 20
           |||||||||| |||||||||
Db     100 GCGCAGCAGGCGCATCGCCA 81


RESULT  31
BE228836/c
LOCUS       BE228836      458 bp    mRNA            EST       07-JUL-2000
DEFINITION  98AS3192 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
            cDNA clone 98AS3192, mRNA sequence.
ACCESSION   BE228836
VERSION     BE228836.1  GI:8955030
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R, Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee
            ,M.C. and Eun,M.Y.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Oryza sativa"
                     /cultivar="Milyang23"
                     /db_xref="taxon:4530"
                     /clone="98AS3192"
                     /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
                     /tissue_type="Immature Seed"
                     /dev_stage="5 days after pollination"
                     /lab_host="E. coli SOLR"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; Directional cDNA library inserted into lambda ZAPII
                     vector  at 5'end with EcoRI and 3' end with Xho I site."
BASE COUNT      117 a    127 c     98 g    103 t     13 others
ORIGIN


  Query Match             73.6%;  Score 18.4;  DB 10;  Length 458;
  Best Local Similarity   86.4%;  Pred. No. 1.9e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 gcagcaggtgcatcgccagcat 24
           |||||  | ||||||||||||||
Db     391 GCAGCGNGNGCATCGCCAGCAT 370


RESULT  32
CNS03HMW/c
LOCUS       CNS03HMW      786 bp    DNA             GSS       17-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            027C16 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL244481
VERSION     AL244481.1  GI:7965493
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
```

```
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 786)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 786)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .786
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="027C16"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG027BB08LP1~end : T7"
BASE COUNT      236 a    188 c    199 g    160 t      3 others
ORIGIN
```

```
  Query Match             73.6%;  Score 18.4;  DB 13;  Length 786;
  Best Local Similarity   95.0%;  Pred. No. 2e+03;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 agcaggtgcatcgccagcat 24
           ||||| ||||||||||||||
Db     496 AGCAGATGCATCGCCAGCAT 477
```

```
RESULT  33
AI901905
LOCUS       AI901905      165 bp    mRNA            EST       27-JUL-1999
DEFINITION  618012B05.x1 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA
            sequence.
ACCESSION   AI901905
VERSION     AI901905.1  GI:5608238
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 165)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 618012  row: B  column: 05.
FEATURES             Location/Qualifiers
     source          1. .165
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="618 - Inbred Tassel cDNA Library"
                     /tissue_type="tassel"
                     /dev_stage="tassel length from 0.1 to 2.5 cm"
                     /lab_host="XLOLR"
                     /note="Organ: tassel; Vector: pAD-GAL4-2.1 (Hybrizap);
                     Inbred tassel library from Schmidt lab"
BASE COUNT       39 a     52 c     53 g     20 t      1 others
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 10;  Length 165;
  Best Local Similarity   87.0%;  Pred. No. 2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagca 23
           |||||||||||| || | ||||||
Db     130 GCGCAGCAGGTACAGCCCCAGCA 152
```

```
RESULT  34
AA741641
LOCUS       AA741641      290 bp    mRNA            EST       10-DEC-1998
DEFINITION  LmLv39p3/248A Leishmania major promastigote full length cDNA
            library from early logarithmic stage (day 3) Leishmania major cDNA
            clone 248A 5', mRNA sequence.
ACCESSION   AA741641
VERSION     AA741641.1  GI:2781620
KEYWORDS    EST.
SOURCE      Leishmania major.
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 290)
  AUTHORS   Almeida,R., Sampaio,I., Schneider,H. and Blackwell,J.M.
  TITLE     Analysis of Leishmania major promastigote library from different
            stages of development
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Blackwell JM
            Cambridge Institute for Medical Research
            Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road,
            Cambridge CB2 2XY, UK
            Tel: 01223 336 143
            Fax: 01223 331 206
            Email: jmb37@cus.cam.ac.uk
            PCR PRimers
            FORWARD: GTAAAACGACGGCCAGT
            BACKWARD: GGAAACAGCTATGACCATG
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 290.
FEATURES             Location/Qualifiers
     source          1. .290
                     /organism="Leishmania major"
                     /strain="LV39"
                     /db_xref="taxon:5664"
                     /clone="248A"
                     /clone_lib="Leishmania major promastigote full length cDNA
                     library from early logarithmic stage (day 3)"
                     /cell_type="Promastigote"
                     /note="Vector: Lambda Zap II; Site_1: XhoI; Site_2: NotI"
BASE COUNT       62 a     86 c     77 g     65 t
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 10;  Length 290;
  Best Local Similarity   87.0%;  Pred. No. 2.1e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagca 23
           ||  ||||||||||||| |||||
Db      77 GCAAAGCAGGTGCATCGTCAGCA 99
```

```
RESULT  35
AU095285/c
LOCUS       AU095285      334 bp    mRNA            EST       30-JUN-2000
DEFINITION  AU095285 Rice cDNA from immature leaf including apical meristem
            (under short day condition) Oryza sativa cDNA clone E61108, mRNA
            sequence.
```

```
ACCESSION   AU095285
VERSION     AU095285.1  GI:8857967
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 334)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from immature leaf including apical meristem (2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
            E61108_1A.
FEATURES             Location/Qualifiers
     source          1. .334
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="E61108"
                     /clone_lib="Rice cDNA from immature leaf including apical
                     meristem (under short day condition)"
                     /dev_stage="immature leaf including apical meristem (under
                     short day condition)"
BASE COUNT       60 a     93 c    123 g     56 t      2 others
ORIGIN

  Query Match             72.8%;  Score 18.2;  DB 10;  Length 334;
  Best Local Similarity   87.0%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcgcagcaggtgcatcgccagca 23
          |||||||| |||||||| || ||
Db    148 GCGCAGCACGTGCATCGGCAACA 126


RESULT  36
BF753362/c
LOCUS       BF753362      349 bp    mRNA            EST       10-JAN-2001
DEFINITION  QV0-CT0581-021000-421-c02 CT0581 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF753362
VERSION     BF753362.1  GI:12079947
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV0&t2=QV0-CT0581-
            021000-421-c02&t3=2000-10-02&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 4
            High quality sequence stop: 349.
FEATURES             Location/Qualifiers
     source          1. .349
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CT0581"
                     /dev_stage="Adult"
                     /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       73 a    106 c     83 g     87 t
ORIGIN

  Query Match             72.8%;  Score 18.2;  DB 11;  Length 349;
  Best Local Similarity   87.0%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcgcagcaggtgcatcgccagca 23
          |||||||| | ||||||||||
Db    235 GCGCAGCAGCAGTATCGCCAGCA 213


RESULT  37
BG268511
LOCUS       BG268511      349 bp    mRNA            EST       20-FEB-2001
DEFINITION  1000201E11.x4 1000 - Unigene I from Maize Genome Project Zea mays
            cDNA, mRNA sequence.
ACCESSION   BG268511
VERSION     BG268511.1  GI:12973499
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 1000201  row: E  column: 11.
FEATURES             Location/Qualifiers
     source          1. .349
                     /organism="Zea mays"
                     /db_xref="dbEST:618012B05.x1"
                     /db_xref="taxon:4577"
                     /clone_lib="1000 - Unigene I from Maize Genome Project"
                     /note="This library represents the unique ESTs found in
                     the first round of EST sequencing at Stanford University
                     for the maize genome project. Sequences are present from
                     libraries 486, 487, 496, 603, 605, 606, 614, 618, 660, 683
                     , 687, 707, and 945. Contigs were assembled using TIGR's
                     CAP program and a representative EST from each contig was
                     selected for the Unigene set. All singlets were also
```

```
                    selected."
BASE COUNT       85 a     87 c     92 g     83 t      2 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 11;  Length 349;
  Best Local Similarity   87.0%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagca 23
           ||||||||||| || | ||||||
Db     269 GCGCAGCAGGTACAGCCCCAGCA 291


RESULT  38
BE397105
LOCUS       BE397105      355 bp    mRNA            EST       21-JUL-2000
DEFINITION  601289536F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3620001 5',
            mRNA sequence.
ACCESSION   BE397105
VERSION     BE397105.1  GI:9342470
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 355)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM291  row: g  column: 10
            High quality sequence stop: 355.
FEATURES             Location/Qualifiers
     source          1. .355
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3620001"
                     /clone_lib="NIH_MGC_8"
                     /tissue_type="Burkitt lymphoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT       94 a     89 c     96 g     76 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 355;
  Best Local Similarity   87.0%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagca 23
           |||||||||  | ||||||||||
Db     112 GCGCAGCAGCAGTATCGCCAGCA 134


RESULT  39
AW952110
LOCUS       AW952110      405 bp    mRNA            EST       01-JUN-2000
DEFINITION  EST364180 MAGE resequences, MAGB Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW952110
VERSION     AW952110.1  GI:8141792
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 35
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1. .405
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGB"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      103 a    105 c    105 g     92 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 405;
  Best Local Similarity   87.0%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagca 23
           |||||||||  | ||||||||||
Db     112 GCGCAGCAGCAGTATCGCCAGCA 134


RESULT  40
AI546214
LOCUS       AI546214      407 bp    mRNA            EST       19-APR-2001
DEFINITION  LD47729.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD47729 5prime similar to X75498:
            D.melanogaster gene encoding proteinA, mRNA sequence.
ACCESSION   AI546214
VERSION     AI546214.1  GI:4463587
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 477  row: C  column: 5
            High quality sequence stop: 333.
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD47729"
```

/clone_lib="LD Drosophila melanogaster embryo pOT2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XL1 Blue"
/note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. "

BASE COUNT 87 a 141 c 108 g 71 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 407;
Best Local Similarity 87.0%; Pred. No. 2.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gcgcagcaggtgcatcgccagca 23
|| |||| |||||| |||||||||
Db 375 GCTCAGCGGGTGCAGCGCCAGCA 397

RESULT 41
AA387004/c
LOCUS AA387004 420 bp mRNA EST 23-APR-1997
DEFINITION vc20g03.r1 Ko mouse embryo 11 5dpc Mus musculus cDNA clone IMAGE:775156 5' similar to TR:G1054889 G1054889 HMGI-C CHIMERIC TRANSCRIPT ;, mRNA sequence.
ACCESSION AA387004
VERSION AA387004.1 GI:2040009
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 420)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:468012
High quality sequence stop: 416.

FEATURES Location/Qualifiers
source 1. .420
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:775156"
/clone_lib="Ko mouse embryo 11 5dpc"
/sex="pooled"
/tissue_type="embryo"
/dev_stage="11.5dpc"
/lab_host="DH10B"
/note="Organ: embryo; Vector: pSPORT1; Site_1: SalI; Site_2: NotI; Total RNAs were extracted from 11.5 dpc embryos (excluding placenta and yolk sac). The double-stranded cDNA was synthesized with an oligo (dT)-1 primer GAGAGAGACTAGTTCTAGATCGCGAGCGGCCGCTTTTTTTTTTTTTTTT 3'. The cDNAs were ligated to LL-Sal3A: 5' GCTATTGACGTCGACTATCC 3' and LL-Sal3B: 5' GGATAGTCGACGTCAAT 3'. The cDNAs were size-selected and amplified by long-range PCR using Ex Taq polymerase for 18 cycles. The PCR-amplifiable cDNA mixture went through one round of equalization and was digested with SalI/NotI and cloned into the SalI/NotI sites of the pSPORT1 plasmid vector (Life Technologies). The library was constructed by Dr. Minoru S. H. Ko and Dr. Xiaohong Wang."

BASE COUNT 75 a 125 c 130 g 90 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 420;
Best Local Similarity 87.0%; Pred. No. 2.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gcgcagcaggtgcatcgccagca 23
|||| |||||||||||||| |||
Db 134 GCGCCTCAGGTGCATCGCCGGCA 112

RESULT 42
BF587149
LOCUS BF587149 436 bp mRNA EST 12-DEC-2000
DEFINITION FM1_32_G07.b1_A003 Floral-Induced Meristem 1 (FM1) Sorghum propinquum cDNA, mRNA sequence.
ACCESSION BF587149
VERSION BF587149.1 GI:11679473
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 436)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt ,L.H.
TITLE An EST database from Sorghum: floral-induced meristems
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 403
POLYA=No.

FEATURES Location/Qualifiers
source 1. .436
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Floral-Induced Meristem 1 (FM1)"
/note="Organ: Floral-induced meristems; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 65 a 156 c 162 g 53 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 436;
Best Local Similarity 87.0%; Pred. No. 2.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gcgcagcaggtgcatcgccagca 23

```
              ||||||||| |||| | ||||||
Db      99 GCGCAGCAGCTGCAGCACCAGCA 121


RESULT  43
BF994016
LOCUS       BF994016      460 bp    mRNA            EST       23-JAN-2001
DEFINITION  MR0-GN0131-021100-103-b06 GN0131 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF994016
VERSION     BF994016.1  GI:12400339
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR0&t2=MR0-GN0131-
            021100-103-b06&t3=2000-11-02&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 8
            High quality sequence stop: 460.
FEATURES              Location/Qualifiers
     source           1. .460
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="GN0131"
                      /dev_stage="Adult"
                      /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                      ; Site_2: SmaI; A mini-library was made by cloning
                      products derived from  ORESTES PCR (U.S. Letters Patent
                      application No. 196,716 -  Ludwig Institute for Cancer
                      Research) profiles into  the pUC 18 vector. Reverse
                      transcription of tissue mRNA and cDNA amplification were
                      performed under low stringency conditions."
BASE COUNT      118 a      89 c     141 g     112 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 11;  Length 460;
  Best Local Similarity   87.0%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 gcagcaggtgcatcgccagcatt 25
           | |||||| |||||||| |||||
Db     317 GTAGCAGGAGCATCGCCTGCATT 339


RESULT  44
BE268250
LOCUS       BE268250      470 bp    mRNA            EST       13-JUL-2000
DEFINITION  601125693F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3345503 5',
            mRNA sequence.
ACCESSION   BE268250
VERSION     BE268250.1  GI:9141852
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM133  row: i  column: 24
            High quality sequence stop: 470.
FEATURES              Location/Qualifiers
     source           1. .470
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3345503"
                      /clone_lib="NIH_MGC_8"
                      /tissue_type="Burkitt lymphoma"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
                      EcoRI; cDNA made by oligo-dT priming.  Directionally
                      cloned into EcoRI/XhoI sites using the following 5'
                      adaptor: GGCACGAG(G).  Size-selected >500bp for average
                      insert size 1.8kb.  Library constructed by Ling Hong in
                      the laboratory of Gerald M. Rubin (University of
                      California, Berkeley) using  ZAP-cDNA synthesis kit
                      (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      119 a      129 c     121 g     101 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 470;
  Best Local Similarity   87.0%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagca 23
           ||||||||| | ||||||||||
Db     227 GCGCAGCAGCAGTATCGCCAGCA 249


RESULT  45
AI542824
LOCUS       AI542824      489 bp    mRNA            EST       19-APR-2001
DEFINITION  SD09292.5prime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD09292 5prime similar to
            X75498: D.melanogaster gene encoding proteinA, mRNA sequence.
ACCESSION   AI542824
VERSION     AI542824.1  GI:4460197
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
```

```
           Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
           Plate: 92  row: H  column: 8
           High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="SD09292"
                     /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                     culture pOT2"
                     /lab_host="DH5-alpha"
                     /note="Vector: pOT2; Site_1: EcoRI; Site_2: Xhol; Sized
                     fractionated cDNAs were directly ligated into pOT2.
                     Plasmid cDNA library."
BASE COUNT      109 a    164 c    129 g     87 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 489;
  Best Local Similarity   87.0%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcgcagcaggtgcatcgccagca 23
           || |||| |||||| ||||||||
Db     427 GCTCAGCGGGTGCAGCGCCAGCA 449


Search completed: December 26, 2001, 12:11:42
Job time: 6464 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:56:58 ; Search time 1755.64 Seconds
(without alignments)
234.917 Million cell updates/sec

Title: US-09-396-196F-6
Perfect score: 25
Sequence: 1 gcaggtgcatcgccagcatttcgat 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	6	AR029499	AR029499 Sequence
2	25	100.0	1041	6	AR034916	AR034916 Sequence
3	25	100.0	1084	6	A11530	A11530 BioB gene o
4	25	100.0	1121	6	E00893	E00893 Genomic DNA
5	25	100.0	5793	1	ECOBIO	J04423 E.coli 7,8-
6	25	100.0	5872	6	A38246	A38246 Sequence 1
7	25	100.0	5872	6	A38251	A38251 Sequence 6
8	25	100.0	5872	6	A93674	A93674 Sequence 1
9	25	100.0	5872	6	A93679	A93679 Sequence 6
10	25	100.0	5872	6	AR101809	AR101809 Sequence
11	25	100.0	5872	6	AR101810	AR101810 Sequence
12	25	100.0	11022	1	AE000180	AE000180 Escherich
13	25	100.0	13501	1	AE005258	AE005258 Escherich
14	25	100.0	297816	1	AP002553	AP002553 Escherich
15	23.4	93.6	5526	1	AF250776	AF250776 Unculture
16	20.8	83.2	965	1	AF250770	AF250770 Unculture
17	19.4	77.6	10593	1	AE003860	AE003860 Xylella f
c 18	19.2	76.8	193351	2	AC068669	AC068669 Homo sapi
c 19	19	76.0	3603	1	PSEHRPIA	L11582 Pseudomonas
20	18.8	75.2	598	11	AF344071	AF344071 Oryza sat
c 21	18.8	75.2	3078	1	AY037869	AY037869 Pseudomon
c 22	18.8	75.2	3148	1	AF192795	AF192795 Pseudomon
23	18.8	75.2	134058	8	AC037425	AC037425 Oryza sat
24	18.6	74.4	11639	1	AE000198	AE000198 Escherich
25	18.6	74.4	11641	1	AE005286	AE005286 Escherich
26	18.6	74.4	12438	1	AE004044	AE004044 Xylella f
27	18.6	74.4	13832	1	D90734	D90734 Escherichia
28	18.6	74.4	200631	2	AC026813	AC026813 Mus muscu
29	18.6	74.4	212908	2	AC068663	AC068663 Mus muscu
30	18.6	74.4	327773	1	AP002554	AP002554 Escherich
31	18.6	74.4	347750	1	AP002998	AP002998 Mesorhizo
c 32	18.4	73.6	119797	2	AC013236	AC013236 Drosophil
c 33	18.4	73.6	128626	3	AC092245	AC092245 Drosophil
34	18.4	73.6	264853	3	AE003644	AE003644 Drosophil
35	18.4	73.6	300962	3	DROSADH04	AE003410 Drosophil
36	18.2	72.8	10864	1	AE004631	AE004631 Pseudomon
c 37	18.2	72.8	14493	1	AE007132	AE007132 Mycobacte
c 38	18.2	72.8	25404	1	D85415	D85415 Pseudomonas
c 39	18.2	72.8	31859	1	MTCY22D7	Z83866 Mycobacteri
c 40	18.2	72.8	103920	2	AC014955	AC014955 Drosophil
c 41	18.2	72.8	104000	2	AC091510	AC091510 Leishmani
c 42	18.2	72.8	281869	3	AE003544	AE003544 Drosophil
43	18	72.0	2923	1	MAV250020	AJ250020 Mycobacte
44	17.8	71.2	479	11	G01446	G01446 Dm0466 Dros
45	17.8	71.2	3748	1	MSU31280	U31280 Methylobaci
46	17.8	71.2	3748	12	CVU39574	U39574 Expression
47	17.8	71.2	35928	3	CELF41H10	U61954 Caenorhabdi
c 48	17.8	71.2	53067	2	AC019940	AC019940 Drosophil
c 49	17.8	71.2	68419	2	AC016014	AC016014 Homo sapi
50	17.8	71.2	110000	2	AL359456_2	Continuation (3 of
51	17.8	71.2	112428	2	AP003906	AP003906 Oryza sat
c 52	17.8	71.2	140156	2	AC017381	AC017381 Drosophil
53	17.8	71.2	148900	2	AC025404	AC025404 Homo sapi
54	17.8	71.2	159506	9	HS34I8	AL021918 Human DNA
55	17.8	71.2	160089	3	AC007813	AC007813 Drosophil
56	17.8	71.2	176687	2	AL365213	AL365213 Homo sapi
57	17.8	71.2	185200	3	AC092401	AC092401 Drosophil
58	17.8	71.2	186989	3	AC007812	AC007812 Drosophil
59	17.8	71.2	191558	3	AC013431	AC013431 Drosophil
60	17.8	71.2	255624	3	AE003723	AE003723 Drosophil
c 61	17.8	71.2	307551	1	AP003015	AP003015 Mesorhizo
62	17.8	71.2	327446	3	AE003500	AE003500 Drosophil
c 63	17.6	70.4	1643	1	D82818	D82818 Pseudomonas
c 64	17.6	70.4	1764	1	AB006480	AB006480 Comamonas
65	17.6	70.4	2740	8	YSACACHS1A	D43627 Candida alb
66	17.6	70.4	3084	6	AR049613	AR049613 Sequence
67	17.6	70.4	3385	8	CACHS1	X52420 C.albicans
68	17.6	70.4	3587	8	D86122	D86122 Arabidopsis
69	17.6	70.4	3617	22	E11448	E11448 cDNA encodi
70	17.6	70.4	7401	3	AF247761	AF247761 Drosophil

c	71	17.6	70.4	11553	1	AE008113
c	72	17.6	70.4	14379	1	AE008283
	73	17.6	70.4	42565	8	CAC35A5
	74	17.6	70.4	45333	9	HS10618A
c	75	17.6	70.4	45453	8	AB022212
c	76	17.6	70.4	45519	1	FVBPOAD2A
c	77	17.6	70.4	61964	3	AC004341
	78	17.6	70.4	72302	2	AC019988
c	79	17.6	70.4	74189	8	AP003225
c	80	17.6	70.4	80954	2	AC011276
	81	17.6	70.4	81487	9	HSJ770C23
	82	17.6	70.4	84961	2	AC014810
	83	17.6	70.4	90788	2	AC005442
c	84	17.6	70.4	102008	2	AP004026
c	85	17.6	70.4	103550	2	AP004003
	86	17.6	70.4	105559	2	AP003989
c	87	17.6	70.4	113328	2	AP003999
	88	17.6	70.4	124927	2	AC010475
	89	17.6	70.4	125439	9	AC010485
	90	17.6	70.4	127653	2	AC036102
c	91	17.6	70.4	159738	2	AC055855
c	92	17.6	70.4	166486	9	CNS01DWC
c	93	17.6	70.4	170186	2	AC027236
c	94	17.6	70.4	171045	2	AL450322
	95	17.6	70.4	172047	2	AC011197
	96	17.6	70.4	179464	2	AC007616
	97	17.6	70.4	183813	2	AC012204
c	98	17.6	70.4	187833	9	AL513303
	99	17.6	70.4	195814	2	AL445308
	100	17.6	70.4	201235	2	AC023156

AE008113	Agrobacte
AE008283	Agrobacte
AL033396	C.albican
AL049773	Human DNA
AB022212	Arabidops
D26094	Flavobacter
AC004341	Drosophil
AC019988	Drosophil
AP003225	Oryza sat
AC011276	Homo sapi
AL118509	Human DNA
AC014810	Drosophil
AC005442	Drosophil
AP004026	Oryza sat
AP004003	Oryza sat
AP003989	Oryza sat
AP003999	Oryza sat
AC010475	Homo sapi
AC010485	Homo sapi
AC036102	Homo sapi
AC055855	Homo sapi
AL137100	Human chr
AC027236	Homo sapi
AL450322	Homo sapi
AC011197	Homo sapi
AC007616	Homo sapi
AC012204	Homo sapi
AL513303	Human DNA
AL445308	Homo sapi
AC023156	Homo sapi

ALIGNMENTS

```
RESULT   1
AR029499
LOCUS        AR029499      1041 bp    DNA              PAT       29-SEP-1999
DEFINITION   Sequence 7 from patent US 5859335.
ACCESSION    AR029499
VERSION      AR029499.1  GI:5941472
KEYWORDS     .
SOURCE       Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE    1  (bases 1 to 1041)
  AUTHORS    Patton,D.Andrew.
  TITLE      Enhanced biotin biosynthesis in plant tissue
  JOURNAL    Patent: US 5859335-A 7 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT       262 a    273 c    305 g    201 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcgat 25
          |||||||||||||||||||||||||
Db     84 GCAGGTGCATCGCCAGCATTTCGAT 108


RESULT   2
AR034916
LOCUS        AR034916      1041 bp    DNA              PAT       29-SEP-1999
DEFINITION   Sequence 7 from patent US 5869719.
ACCESSION    AR034916
VERSION      AR034916.1  GI:5950521
KEYWORDS     .
SOURCE       Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE    1  (bases 1 to 1041)
  AUTHORS    Patton,D.A.
  TITLE      Transgenic plants having increased biotin content
  JOURNAL    Patent: US 5869719-A 7 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT       262 a    273 c    305 g    201 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcgat 25
          |||||||||||||||||||||||||
Db     84 GCAGGTGCATCGCCAGCATTTCGAT 108


RESULT   3
A11530
LOCUS        A11530        1084 bp    DNA              PAT       09-FEB-1994
DEFINITION   BioB gene of E.coli with primers.
ACCESSION    A11530
VERSION      A11530.1  GI:490218
KEYWORDS     .
SOURCE       Escherichia coli.
  ORGANISM   Escherichia coli
             Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
             Escherichia.
REFERENCE    1  (bases 1 to 1084)
  AUTHORS    .
  JOURNAL    Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            24. .1064
                     /gene="BIOB"
     CDS             24. .1064
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490219"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
BASE COUNT       271 a    286 c    318 g    209 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcgat 25
          |||||||||||||||||||||||||
Db    107 GCAGGTGCATCGCCAGCATTTCGAT 131


RESULT   4
E00893
LOCUS        E00893        1121 bp    DNA              PAT       29-SEP-1997
DEFINITION   Genomic DNA encoding biotin Synthetase.
```

```
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES              Location/Qualifiers
     source           1. .1121
                      /organism="Escherichia coli"
                      /db_xref="taxon:562"
BASE COUNT      289 a    296 c    325 g    211 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcgat 25
          |||||||||||||||||||||||||
Db    125 GCAGGTGCATCGCCAGCATTTCGAT 149


RESULT   5
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES              Location/Qualifiers
     source           1. .5793
                      /organism="Escherichia coli"
                      /strain="K-12"
                      /db_xref="taxon:562"
     CDS              complement(98. .574)
                      /note="ORF 1"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="AAA23513.1"
                      /db_xref="GI:455168"
                      /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                      SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                      YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene             complement(633. .1925)
                      /gene="bioA"
     CDS              complement(633. .1925)
                      /gene="bioA"
                      /codon_start=1
                      /transl_table=11
                      /product="7,8-diamino-pelargonic acid aminotransferase"
                      /protein_id="AAA23514.1"
                      /db_xref="GI:457106"
                      /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                      LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                      ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                      SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                      GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                      LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                      QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                      PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene             2012. .3052
                      /gene="bioB"
     CDS              2012. .3052
                      /gene="bioB"
                      /codon_start=1
                      /transl_table=11
                      /product="biotin synthetase"
                      /protein_id="AAA23515.1"
                      /db_xref="GI:145425"
                      /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                      LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                      KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                      YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                      VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                      FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                      PDTDEYYNAAAL"
     gene             3049. .4203
                      /gene="bioF"
     CDS              3049. .4203
                      /gene="bioF"
                      /codon_start=1
                      /transl_table=11
                      /product="7-keto-8-amino pelargonic acid synthetase"
                      /protein_id="AAA23516.1"
                      /db_xref="GI:145426"
                      /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                      SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                      LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                      LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                      SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                      RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                      AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene             4190. .4945
                      /gene="bioC"
     CDS              4190. .4945
                      /gene="bioC"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="AAA23517.1"
                      /db_xref="GI:145427"
                      /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                      HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                      TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
```

```
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4938. .5597
                     /gene="bioD"
     CDS             4938. .5597
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAA23518.1"
                     /db_xref="GI:145428"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                     T"
BASE COUNT     1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match             100.0%;  Score 25;  DB 1;  Length 5793;
  Best Local Similarity   100.0%;  Pred. No. 0.44;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||||||||||||||||||||
Db    2095 GCAGGTGCATCGCCAGCATTTCGAT 2119


RESULT   6
A38246
LOCUS       A38246       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /number=1
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAA02325.1"
                     /db_xref="GI:2294845"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=3
                     /product="PROTEIN"
                     /protein_id="CAA02326.1"
                     /db_xref="GI:2294846"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:
                     8-AMINO-7-OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /number=5
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAA02327.1"
                     /db_xref="GI:2294847"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=6
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAA02328.1"
                     /db_xref="GI:2294848"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNIPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
```

```
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.44;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcgat 25
          |||||||||||||||||||||||||
Db    200 GCAGGTGCATCGCCAGCATTTCGAT 224


RESULT   7
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.44;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcgat 25
          |||||||||||||||||||||||||
Db    200 GCAGGTGCATCGCCAGCATTTCGAT 224


RESULT   8
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
```

```
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORF1 RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.44;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||||||||||||||||||||
Db     200 GCAGGTGCATCGCCAGCATTTCGAT 224


RESULT   9
A93679
LOCUS       A93679       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.44;
```

```
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||||||||||||||||||||
Db     200 GCAGGTGCATCGCCAGCATTTCGAT 224


RESULT  10
AR101809
LOCUS       AR101809     5872 bp    DNA              PAT       14-FEB-2001
DEFINITION  Sequence 1 from patent US 6083712.
ACCESSION   AR101809
VERSION     AR101809.1  GI:12812607
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.44;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||||||||||||||||||||
Db     200 GCAGGTGCATCGCCAGCATTTCGAT 224


RESULT  11
AR101810
LOCUS       AR101810     5872 bp    DNA              PAT       14-FEB-2001
DEFINITION  Sequence 6 from patent US 6083712.
ACCESSION   AR101810
VERSION     AR101810.1  GI:12812608
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.44;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||||||||||||||||||||
Db     200 GCAGGTGCATCGCCAGCATTTCGAT 224


RESULT  12
AE000180
LOCUS       AE000180    11022 bp    DNA              BCT       01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 70 of 400 of the complete
            genome.
ACCESSION   AE000180 U00096
VERSION     AE000180.1  GI:1786988
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 11022)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 11022)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data.  Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .11022
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   <1. .130
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 4 REP sequences"
     gene            complement(147. .1430)
                     /gene="ybhC"
                     /note="b0772"
```

```
CDS             complement(147. .1430)
                /gene="ybhC"
                /function="orf; Not classified"
                /note="f427; 98 pct identical to fragment YBHC_ECOLI
                SW:P46130 (300 aa) but contains 127 additional C-terminal
                residues"
                /codon_start=1
                /transl_table=11
                /product="putative pectinesterase"
                /protein_id="AAC73859.1"
                /db_xref="GI:1786989"
                /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPAGTQGVTHTTIQAAVD
                AAIIKRTNKRQYIAVMPGEYQGTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
                NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSNIYYGFLAVNSRFN
                AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
promoter        complement(1507. .1535)
                /note="factor Sigma70; predicted +1 start at 806574"
gene            complement(1582. .2058)
                /gene="ybhB"
                /note="b0773"
CDS             complement(1582. .2058)
                /gene="ybhB"
                /function="orf; Unknown"
                /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC73860.1"
                /db_xref="GI:1786990"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
protein_bind    complement(2108. .2124)
                /note="central position to predicted promoter:85.5"
                /bound_moiety="RhaS predicted site"
gene            complement(2117. .3406)
                /gene="bioA"
                /note="b0774"
CDS             complement(2117. .3406)
                /gene="bioA"
                /EC_number="2.6.1.62"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                /codon_start=1
                /transl_table=11
                /product="7,8-diaminopelargonic acid synthetase"
                /protein_id="AAC73861.1"
                /db_xref="GI:1786991"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
promoter        complement(2193. .2221)
                /gene="bioA"
                /note="factor Sigma70; predicted +1 start at 807260"
protein_bind    3411. .3450
                /note="central position to bioB promoter: -20"
                /bound_moiety="BioB documented site"
protein_bind    3411. .3450
                /note="central position to predicted promoter: -20;genetic
                evidence for the site"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to predicted promoter:50"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to bioA promoter:50"
                /bound_moiety="BioB documented site"
promoter        3413. .3441
                /note="factor Sigma70; promoter bioB; documented +1 at
                808525"
promoter        complement(3447. .3473)
                /note="factor Sigma70; promoter bioA; documented +1 at
                808515"
gene            3493. .4533
                /gene="bioB"
                /note="b0775"
CDS             3493. .4533
                /gene="bioB"
                /EC_number="2.8.1.-"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                /codon_start=1
                /transl_table=11
                /product="biotin synthesis, sulfur insertion?"
                /protein_id="AAC73862.1"
                /db_xref="GI:1786992"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            4530. .5684
                /gene="bioF"
                /note="b0776"
CDS             4530. .5684
                /gene="bioF"
                /EC_number="2.3.1.47"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                /codon_start=1
                /transl_table=11
                /product="8-amino-7-oxononanoate synthase"
                /protein_id="AAC73863.1"
                /db_xref="GI:1786993"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
```

```
  Query Match             100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity   100.0%;  Pred. No. 0.44;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||||||||||||||||||||
Db    3576 GCAGGTGCATCGCCAGCATTTCGAT 3600


RESULT  13
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
```

AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL Nature 409 (6819), 529-533 (2001)
MEDLINE 21074935
PUBMED 11206551
REFERENCE 2 (bases 1 to 13501)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES Location/Qualifiers

source 1. .13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"

misc_feature <1. .7576
/note="O-island #36; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655; Cryptic prophage CP-933K; includes one copy of the 13 bp direct repeat that flanks the prophage"

gene 66. .665
/gene="lomK"
/note="Z0981"

CDS 66. .665
/gene="lomK"
/function="putative membrane; Other or unknown (Phage or Prophage Related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to residues 1 to 199 of 199 from GenPept 118 : gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom [prophage P-EibA]"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein of prophage CP-933K"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MRKVCAAILSAAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG PSVRVNEWFSAYAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"

gene 719. .2041
/gene="Z0982"

CDS 719. .2041
/gene="Z0982"
/function="putative structure; Structural component (Phage or Prophage Related)"
/note="Residues 164 to 440 of 440 are 68.79 pct identical to residues 381 to 645 of 645 from GenPept 118 : gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative tail fiber protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="putative tail component of prophage CP-933K"
/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQPGTLNDFLGAMTEDDAR PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE SARQAAESAASAKKSEEASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA GPQGPKGETGAAGPVGATGPQGPKGDPGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"

gene 2004. .2312
/gene="Z0984"

CDS 2004. .2312
/gene="Z0984"
/function="orf; Other or unknown (Phage or Prophage Related)"
/note="Residues 1 to 102 of 102 are 98.03 pct identical to residues 28 to 129 of 129 from GenPept 118 : gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520) hypothetical protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK "

gene 2489. .3469
/gene="Z0985"

CDS 2489. .3469
/gene="Z0985"
/function="orf; Other or unknown (Phage or Prophage Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical to residues 18 to 331 of 336 from GenPept 118 : gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"

gene 3530. .4522
/gene="Z0986"

CDS 3530. .4522
/gene="Z0986"
/function="orf; Other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"

gene 5350. .6231
/gene="Z0989"

CDS 5350. .6231
/gene="Z0989"
/function="orf; Other or unknown (Phage or Prophage Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV

```
                     KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGEISRKIEVMLNASPVSAAM
                     ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                     FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDIIGIRMDKINGESLLNISSLPAQ
                     AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                     SYHGGKQDLISVVLSKI"
     gene            6462. .7160
                     /gene="Z0990"
     CDS             6462. .7160
                     /gene="Z0990"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55143.1"
                     /db_xref="GI:12513758"
                     /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLRF
                     QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                     IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                     SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                     VSLGFQQVRLHPLL"
     gene            complement(7668. .8144)
                     /gene="ybhB"
                     /note="Z0992"
     CDS             complement(7668. .8144)
                     /gene="ybhB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                     residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                     MG1655: B0773"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55144.1"
                     /db_xref="GI:12513759"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(8203. .9492)
                     /gene="bioA"
                     /note="Z0993"
     CDS             complement(8203. .9492)

  Query Match          100.0%;  Score 25;  DB 1;  Length 13501;
  Best Local Similarity  100.0%;  Pred. No. 0.43;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||||||||||||||||||||
Db    9662 GCAGGTGCATCGCCAGCATTTCGAT 9686


RESULT  14
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .1245
                     /gene="ECs0753"
     CDS             79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
     gene            1245. .2114
                     /gene="ECs0754"
     CDS             1245. .2114
                     /gene="ECs0754"
                     /note="similar to SUCD_ECOLI gi|1786949 percent identity
                     100 in 289 aa (Conserved in E.coli K-12)"
                     /codon_start=1
```

/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
LADIGEALKTVLK"

gene 2781. .3686
/gene="ECs0755"

CDS 2781. .3686
/gene="ECs0755"
/note="probable transcriptional regulator,similar to transcriptional regulators e.g. glycine cleavage system transcription activator (qcv operon activator) - Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
EKSDSPHLQQFREWLLAKSVVPQEM"

gene complement(3720. .4322)
/gene="ECs0756"

CDS complement(3720. .4322)
/gene="ECs0756"
/note="probable cob(I)alamin adenosyltransferase,similar to cob(I)alamin adenosyltransferases (corrinoid adenosyltransferases) e.g. [Escherichia coli] gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative cob(I)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"

gene complement(4332. .5984)
/gene="ECs0757"

CDS complement(4332. .5984)
/gene="ECs0757"
/note="probable fumarate hydratase,similar to fumarate hydratases e.g. fumarate hydratase class I, aerobic (fumarase) - Escherichia coli gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSIGGAAALLAQEYVKSLRCLEYPELG
MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"

gene complement(6092. .7372)
/gene="ECs0758"

CDS complement(6092. .7372)
/gene="ECs0758"
/note="probable transport protein,similar to glutamate/aspartate transport proteins (proton glutamate symport proteins) e.g. [Bacillus stearothermophilus] gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416 aa,also similar to C4-dicarboxylate transpor"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"

gene complement(7533. .7850)
/gene="ECs0759"

CDS complement(7533. .7850)
/gene="ECs0759"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

Query Match 100.0%; Score 25; DB 1; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcaggtgcatcgccagcatttcgat 25
||||||||||||||||||||||||||
Db 92394 GCAGGTGCATCGCCAGCATTTCGAT 92418

RESULT 15
AF250776
LOCUS AF250776 5526 bp DNA BCT 31-JAN-2001
DEFINITION Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA), biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and biotin biosynthesis protein BioC (bioC) genes, complete cds; and dethiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION AF250776
VERSION AF250776.1 GI:12620124
KEYWORDS .
SOURCE uncultured bacterium pCosHE2.
ORGANISM uncultured bacterium pCosHE2
Bacteria; environmental samples.
REFERENCE 1 (bases 1 to 5526)
AUTHORS Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
TITLE Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia
JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
MEDLINE 20575196
PUBMED 11133432
REFERENCE 2 (bases 1 to 5526)
AUTHORS Entcheva,P., Liebl,W. and Streit,W.R.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES Location/Qualifiers

```
     source          1. .5526
                     /organism="uncultured bacterium pCosHE2"
                     /db_xref="taxon:143797"
                     /clone="pCosHE2"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     CDS             complement(52. .528)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical 17.1 kDa protein in modC-bioA
                     intergenic region"
                     /protein_id="AAG60577.1"
                     /db_xref="GI:12620125"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKAG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASCAMVGFNVHFHSLASASITAMFS"
     gene            complement(587. .1876)
                     /gene="bioA"
     CDS             complement(587. .1876)
                     /gene="bioA"
                     /note="7,8-diaminopelargonic acid
                     synthetase-aminotransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="DAPA-aminotransferase BioA"
                     /protein_id="AAG60578.1"
                     /db_xref="GI:12620126"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LKCVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            1963. .3003
                     /gene="bioB"
     CDS             1963. .3003
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase BioB"
                     /protein_id="AAG60579.1"
                     /db_xref="GI:12620127"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPNERDMPYLEQMVQGVKDLGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3000. .4154
                     /gene="bioF"
     CDS             3000. .4154
                     /gene="bioF"
                     /note="8-amino-7-oxononanoate synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="KAPA synthetase BioF"
                     /protein_id="AAG60580.1"
                     /db_xref="GI:12620128"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQLVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRRDEGDARREKLVSLIAHFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4141. .4896
                     /gene="bioC"
     CDS             4141. .4896
                     /gene="bioC"
                     /note="reaction step prior to synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin biosynthesis protein BioC"
                     /protein_id="AAG60581.1"
                     /db_xref="GI:12620129"
                     /translation="MATVNKQAIAAAFGRAAALYEQHADLQRQSADVLLAMLPERKYT
                     HVLDAGCGPGWMTRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRSKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4883. .>5526
                     /gene="bioD"
     CDS             4883. .>5526
                     /gene="bioD"
                     /note="DTB synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase BioD"
                     /protein_id="AAG60582.1"
                     /db_xref="GI:12620130"
                     /translation="MLVSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVAS
                     GSEKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESSVMS
                     SGLRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAML
                     TAQAIQHAGLTLAGWVANDVTPPGKRHAEYITTLTRMIPAPLLGEIPWLAENPE"
BASE COUNT     1274 a    1507 c    1567 g    1178 t
ORIGIN


  Query Match             93.6%;  Score 23.4;  DB 1;  Length 5526;
  Best Local Similarity   96.0%;  Pred. No. 2.5;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccaqcatttcgat 25
           |||||| ||||||||||||||||||||
Db    2046 GCAGGTCCATCGCCAGCATTTCGAT 2070


RESULT  16
AF250770
LOCUS       AF250770      965 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosHE1 DAPA-aminotransferase (bioA) and
            biotin synthase (bioB) genes, partial cds.
ACCESSION   AF250770
VERSION     AF250770.1  GI:12620104
KEYWORDS    .
SOURCE      uncultured bacterium pCosHE1.
  ORGANISM  uncultured bacterium pCosHE1
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 965)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 965)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .965
                     /organism="uncultured bacterium pCosHE1"
                     /db_xref="taxon:143796"
                     /clone="pCosHE1"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     gene            complement(<1. .438)
                     /gene="bioA"
     CDS             complement(<1. .438)
                     /gene="bioA"
                     /note="7,8-diaminopelargonic acid
```

```
                     synthetase-aminotransferase; similar to BioA"
                     /codon_start=1
                     /transl_table=11
                     /product="DAPA-aminotransferase"
                     /protein_id="AAG60563.1"
                     /db_xref="GI:12620105"
                     /translation="MTQDDLAFDQQHIWHPYTSMTRPXPVYPVASAXAHCQRRRRLVD
                     GMSSWWAAIHGYNHPRLNAAMKAQIEQMSHVMFDGITHQPAVDLCRRXVAMTPDALEC
                     VFLADSGSVAXEVXDENGAASTAPLPADRPPERYHXAFFIPXLQ"
     gene            525. .>965
                     /gene="bioB"
     CDS             525. .>965
                     /gene="bioB"
                     /note="similar to BioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase"
                     /protein_id="AAG60564.1"
                     /db_xref="GI:12620106"
                     /translation="MXHHARWTMSQVTELFNKPFLELMFEAQQVHRQHFDPRHVQXST
                     LLSIKTGACPEDCKYCPQSARYKTGLESERLMEVEQVLDSARKAKNAGSTRFCMGAAW
                     KKPHTIVTCPYLEQMVKGVKEMGLEALYDPRHAERRAXXSACLLA"
BASE COUNT      230 a    256 c    285 g    180 t     14 others
ORIGIN


  Query Match             83.2%;  Score 20.8;  DB 1;  Length 965;
  Best Local Similarity   91.7%;  Pred. No. 41;
  Matches   22;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcgat 25
           |||||||||| ||||| ||||||
Db     609 CAGGTGCATCGTCAGCACTTCGAT 632


RESULT  17
AE003860
LOCUS       AE003860    10593 bp    DNA             BCT       15-JUN-2001
DEFINITION  Xylella fastidiosa 9a5c, section 6 of 229 of the complete genome.
ACCESSION   AE003860 AE003849
VERSION     AE003860.1  GI:9104830
KEYWORDS    .
SOURCE      Xylella fastidiosa 9a5c.
  ORGANISM  Xylella fastidiosa 9a5c
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xylella.
REFERENCE   1  (bases 1 to 10593)
  AUTHORS   Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
            Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrer,H.,
            Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
            Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H.,
            Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
            Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
            Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
            Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
            and Marino,C.L.
  TITLE     The genome sequence of the plant pathogen Xylella fastidiosa. The
            Xylella fastidiosa Consortium of the Organization for Nucleotide
            Sequencing and Analysis
  JOURNAL   Nature 406 (6792), 151-157 (2000)
  MEDLINE   20365717
   PUBMED   10910347
REFERENCE   2  (bases 1 to 10593)
  AUTHORS   Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
            Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
            Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
            Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
            Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,
            Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
            Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
            Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
            Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
            Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
            Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
            Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
            Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
            Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
            Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
            Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
            Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
            Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
            de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
            Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
            Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
            Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
            Tsuhako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
            Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
            Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
            13083-970, Brazil
FEATURES             Location/Qualifiers
     source          1. .10593
                     /organism="Xylella fastidiosa 9a5c"
                     /db_xref="taxon:160492"
                     /clone="9a5c"
     gene            complement(277. .1356)
                     /gene="XF0061"
     CDS             complement(277. .1356)
                     /gene="XF0061"
                     /note="similar to SP|P07674 (percent identity: 28 %/query
                     alignment coverage:  64.9 %/subject alignment coverage:
                     65.1 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional repressor"
                     /protein_id="AAF82874.1"
                     /db_xref="GI:9104831"
                     /translation="MLLLQTLFELDRATAMSLFGNRIKPTQEQNQLATVAFSGEHIEE
                     IDPELIDLEKQMRSKDNPGFTIESLMELGNDMKRDGQHEPVILRKNSKKPGRYLMVAG
                     ERRWLACKAVDIKLKAVIRELTDDQVRRVQRAENIQRENLTQLEIAVALRDDKERLGT
                     LEKVAAEWNKGINWVSERINFLEVVETNSLTSQAVSEGITADITAINDLYRLEKVNQT
                     AAKTIIDQAKVDPGINMRKTVREKLQNAKPTTNEKNATTDSRMQESKTDFQEGETHLQ
                     NISNQEKPATVASAPLPPDAAQLVALAAKQMETLCTEVSALSKILHQRQDVISALGDV
                     LQNVVQAQAALFRTRHVLASLANSK"
     gene            complement(1331. .1657)
                     /gene="XF0062"
     CDS             complement(1331. .1657)
                     /gene="XF0062"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF82875.1"
                     /db_xref="GI:9104832"
                     /translation="MYHRSIFQRSNSIHNETVKHKCWSDLQNHDIDTTYPLPNSTNNI
                     LEWDQNFKNTNEICAENTSPIQSIYRPRPPEQHVNIFPILIPHLMPVQHLKMYQARHA
                     VVTDIV"
     gene            complement(2427. .3152)
                     /gene="XF0063"
     CDS             complement(2427. .3152)
                     /gene="XF0063"
                     /note="similar to GI|5921548 (percent identity: 37 %/query
                     alignment coverage:  93.4 %/subject alignment coverage:
                     95.7 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="competence protein F"
                     /protein_id="AAF82876.1"
```

```
                     /db_xref="GI:9104833"
                     /translation="MKDAVNSNH1LPDTWCDRLLRWLLPSYCLACNEPGHQGKTLCAI
                     CQLLLPRMENACLRCAAPLKELEGIHPTCGPCLYDPPPLETAHAAFLYHWPINKL1SK
                     FKFHQDLACGHLLADNMAMAAQHWPRPQALVPVSLHHGRLRTRGYDQALELARHVGNR
                     LQLPCLPALRRVRATLPQSQLNANKRRSNLHNAFIPIHPLPAHVALVDDVMTTGTTLH
                     AAARALRHTGVTRIDAW1CARVP"
gene                 3202. .4227
                     /gene="XF0064"
CDS                  3202. .4227
                     /gene="XF0064"
                     /note="similar to SP|P36569 (percent identity: 60 %/query
                     alignment coverage:  94.4 %/subject alignment coverage:
                     93.1 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder/Start codon shift: 66"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase"
                     /protein_id="AAF82877.1"
                     /db_xref="GI:9104834"
                     /translation="MSSIIRYDWTAEELHALFDLSLPELLYRAASVHRQHFDPAEIQV
                     STLLSVKTGGCPEDCSYCPQAQRYDTGVTAQKLMDVDAVVAKARQAKLAGASRFCMGA
                     AWRSPKDRDIPKIAAMIREVKALGLETCATLGMLNTCQAQALKDAGLDYYNHNVDTSP
                     DFYDSVIHTRQYQDRLDTLAHVRDVGLKTCCGGIVGMGETRQHRVGLLLTLATLPAHP
                     DSVPVNLLVQVAGTPLHGTQTLDPFEFVRMIAVARITMPRSMVRLSAGRESMSDEIQA
                     LCFMAGANS1FYGEKLLTTANPETERDQALFQRLGLRPMHLMENVSNQDQHHGNVHAD
                     IACKHVV"
gene                 4281. .4382
                     /gene="XF0065"
CDS                  4281. .4382
                     /gene="XF0065"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF82878.1"
                     /db_xref="GI:9104835"
                     /translation="MATLVDVVSFMYDGVNLDSWSDGGERRKVVEFS"
gene                 4486. .5673
                     /gene="XF0066"
CDS                  4486. .5673
                     /gene="XF0066"
                     /note="similar to SP|O34731 (percent identity: 33 %/query
                     alignment coverage:  63.3 %/subject alignment coverage:
                     96.2 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF82879.1"
                     /db_xref="GI:9104836"
                     /translation="MSLINHKNGVSCNCGSTCSRDILNRLLWRRVVSGCNRYPVHLSY
                     LALSSMCFTFRCSFLATIVVALLTACSSAPRSASRLSSVSQLSDISLPPLRIGIALGG
                     GAAKGFAHIGVLKMLEANGLAPSVVAGTSAGSVVGALYASGMDSFKIQTKAVAMDESN
                     IRDLRFFSGGLVQGQKLQNYVNELVGKRPIEKLNKPFGAIATRLEDGQRTMFVRGNVG
                     EAVRASCSIPGVFEPVTIGAYHYVDGGITSPVPVDAVRQLGADFVIAVDISSKATGKN
                     PGGLVGTVNQSISIMGQHLGEAELKRANIVIHPKVLDIGSTDFDQRNAAILEGERATI
                     AVIPEIREKLAQLQQQRIQAARKTANQSAIEAVESYRRCQNSRTRWEKLRGKEKDCTK
                     SSL"
gene                 6138. .6377
                     /gene="XF0067"
CDS                  6138. .6377
                     /gene="XF0067"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF82880.1"
                     /db_xref="GI:9104837"
                     /translation="MSCKEIFEYIVIKQCWLPMFFAQMGMVMGAIWCADRYPPIVL1S
                     RCVLSYYVGWGYRVRFSVILWPMCYWLGVM1VLAF"
gene                 complement(6494. .7495)
CDS                  complement(6494. .7495)
                     /gene="XF0068"
                     /note="similar to SP|P26601 (percent identity: 52 %/query
                     alignment coverage:  86.8 %/subject alignment coverage:
                     99.7 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hydroxybenzoate octaprenyltransferase"
                     /protein_id="AAF82881.1"
                     /db_xref="GI:9104838"
                     /translation="MILNQEHEHIQVYNHHLNSFQFPLNPSSDNSPVSMAYERFTSAI
                     TLLMHWRNRLDPYWKLARGDRPVGFLLLLWPTWWALWLAAGGVPPWWTLCVFTTGIWL
                     TRSAGCVINDYTDRWLDPHVERTRIRPLVTGAVSPRNALLMFATLMLIAFGLVLTMNQ
                     LTVLLSVVGLFLAMTYPYLKRYTYLPQVYLGLAFGWGIPMAFAAIQGKVPTQAWLLYV
                     ANILWTTAYDTWCAMVDRDDDIKMGAKSTAILFADLDLTVQGVLYTLMLFTLCLVGLR
                     ATLSHTYWISLIAAVALIGYQFIIARRREPTACFRAFMHNNWVGMTIFAGIALATTH"

  Query Match             77.6%;  Score 19.4;  DB 1;  Length 10593;
  Best Local Similarity   95.2%;  Pred. No. 1.8e+02;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 gtgcatcgccagcatttcgat 25
           ||||||| ||||||||||||
Db    3295 GTGCATCGTCAGCATTTCGAT 3315


RESULT  18
AC068669/c
LOCUS       AC068669   193351 bp    DNA              HTG       07-DEC-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-749I16 map 17, WORKING DRAFT
            SEQUENCE, 25 unordered pieces.
ACCESSION   AC068669
VERSION     AC068669.3  GI:11597011
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 193351)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-749I16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 193351)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 7, 2000 this sequence version replaced gi:8779952.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
```

```
-------------- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L10045
    Center clone name: 749_I_16
-------------- Summary Statistics
    Sequencing vector: M13; M77815; 36% of reads
    Sequencing vector: Plasmid; n/a; 64% of reads
    Chemistry: Dye-terminator Big Dye; 100% of reads
    Assembly program: Phrap; version 0.960731
    Consensus quality: 185301 bases at least Q40
    Consensus quality: 188331 bases at least Q30
    Consensus quality: 189713 bases at least Q20
    Insert size: 177000; agarose-fp
    Insert size: 190951; sum-of-contigs
    Quality coverage: 8.3 in Q20 bases; agarose-fp
    Quality coverage: 7.7 in Q20.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1    15647: contig of 15647 bp in length
*    15648 15747: gap of      100 bp
*    15748    16321: contig of 574 bp in length
*    16322 16421: gap of      100 bp
*    16422    17197: contig of 776 bp in length
*    17198 17297: gap of      100 bp
*    17298    17904: contig of 607 bp in length
*    17905 18004: gap of      100 bp
*    18005    19904: contig of 1900 bp in length
*    19905 20004: gap of      100 bp
*    20005    21868: contig of 1864 bp in length
*    21869 21968: gap of      100 bp
*    21969    24760: contig of 2792 bp in length
*    24761 24860: gap of      100 bp
*    24861    27006: contig of 2146 bp in length
*    27007 27106: gap of      100 bp
*    27107    29532: contig of 2426 bp in length
*    29533 29632: gap of      100 bp
*    29633    56088: contig of 26456 bp in length
*    56089 56188: gap of      100 bp
*    56189    59962: contig of 3774 bp in length
*    59963 60062: gap of      100 bp
*    60063    65173: contig of 5111 bp in length
*    65174 65273: gap of      100 bp
*    65274    68681: contig of 3408 bp in length
*    68682 68781: gap of      100 bp
*    68782    73500: contig of 4719 bp in length
*    73501 73600: gap of      100 bp
*    73601    78056: contig of 4456 bp in length
*    78057 78156: gap of      100 bp
*    78157    84041: contig of 5885 bp in length
*    84042 84141: gap of      100 bp
*    84142    93337: contig of 9196 bp in length
*    93338 93437: gap of      100 bp
*    93438   102240: contig of 8803 bp in length
*   102241 102340: gap of      100 bp
*   102341   113166: contig of 10826 bp in length
*   113167 113266: gap of      100 bp
*   113267   123311: contig of 10045 bp in length
*   123312 123411: gap of      100 bp
*   123412   133265: contig of 9854 bp in length
*   133266 133365: gap of      100 bp
*   133366   148098: contig of 14733 bp in length
*   148099 148198: gap of      100 bp
*   148199   165273: contig of 17075 bp in length
*   165274 165373: gap of      100 bp
*   165374   190517: contig of 25144 bp in length
*   190518 190617: gap of      100 bp
*   190618   193351: contig of 2734 bp in length.
FEATURES             Location/Qualifiers
     source          1. .193351
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-749I16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .15647
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    15748. .16321
                     /note="assembly_fragment"
     misc_feature    16422. .17197
                     /note="assembly_fragment"
     misc_feature    17298. .17904
                     /note="assembly_fragment"
     misc_feature    18005. .19904
                     /note="assembly_fragment"
     misc_feature    20005. .21868
                     /note="assembly_fragment"
     misc_feature    21969. .24760
                     /note="assembly_fragment"
     misc_feature    24861. .27006
                     /note="assembly_fragment"
     misc_feature    27107. .29532
                     /note="assembly_fragment"
     misc_feature    29633. .56088
                     /note="assembly_fragment"
     misc_feature    56189. .59962
                     /note="assembly_fragment"
     misc_feature    60063. .65173
                     /note="assembly_fragment"
     misc_feature    65274. .68681
                     /note="assembly_fragment"
     misc_feature    68782. .73500
                     /note="assembly_fragment"
     misc_feature    73601. .78056
                     /note="assembly_fragment"
     misc_feature    78157. .84041
                     /note="assembly_fragment"
     misc_feature    84142. .93337
                     /note="assembly_fragment"
     misc_feature    93438. .102240
                     /note="assembly_fragment"
     misc_feature    102341. .113166
                     /note="assembly_fragment"
     misc_feature    113267. .123311
                     /note="assembly_fragment"
     misc_feature    123412. .133265
                     /note="assembly_fragment"
     misc_feature    133366. .148098
                     /note="assembly_fragment"
     misc_feature    148199. .165273
                     /note="assembly_fragment"
     misc_feature    165374. .190517
                     /note="assembly_fragment"
     misc_feature    190618. .193351
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    48960 a  47312 c  47389 g  47267 t   2423 others
ORIGIN


  Query Match           76.8%;  Score 19.2;  DB 2;  Length 193351;
  Best Local Similarity   87.5%;  Pred. No. 2e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy       2 caggtgcatcgccagcatttcgat 25
           |||| ||||| |||||||||| ||
Db   43485 CAGGGGCATCCCCAGCATTTCTAT 43462


RESULT  19
PSEHRPIA/c
LOCUS       PSEHRPIA     3603 bp    DNA             BCT       21-OCT-1993
DEFINITION  Pseudomonas syringae hrpIJ genes, complete cds.
ACCESSION   L11582
VERSION     L11582.1  GI:151267
KEYWORDS    hrpI gene; protein translocation.
SOURCE      Pseudomonas syringae (strain 61) DNA.
  ORGANISM  Pseudomonas syringae
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 3603)
  AUTHORS   Huang,H.-C., Xiao,Y., Lin,R.-H., Lu,Y., Hutcheson,S.W. and
            Collmer,A.
  TITLE     Characterization of the Pseudomonas syringae pv. syringae 61 hrpJ
            and hrpI genes: homology of HrpI to a superfamily of proteins
            associated with protein translocation
  JOURNAL   Mol. Plant Microb Interact. 6, 515-520 (1993)
FEATURES             Location/Qualifiers
     source          1. .3603
                     /organism="Pseudomonas syringae"
                     /db_xref="taxon:317"
     gene            1330. .3417
                     /gene="hrpI"
     CDS             1330. .3417
                     /gene="hrpI"
                     /note="hypersensitive response and pathogenicity gene,
                     membrane-spanning protein; hrpI superfamily proteins are
                     associated with protein translocation; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="HrpI"
                     /protein_id="AAA03015.1"
                     /db_xref="GI:151268"
                     /translation="MNRVINFLNMVALSAMRRSELVGAFFVIAIVFMMITPLPTGLID
                     VLIAVNICISCLLIMLAMHLPRPLAFSTFPAVLLLTTMFRLALSVSTTRLILLNQDAG
                     HIVEAFGQFVVGGNLAVGLVIFLILTVVNFLVITKGSERVAEVGARFTLDAMPGKQMS
                     IDSDLRANLITVHEARKRRAELNKESQLFGAMDGAMKFVNGDAIASLIIVAINMIGGI
                     SIGVLQHNMAAGDALQLYTVLTIGDGLIAQIPALLISVTSGMIITRVPNTEAGVEANI
                     GREIAEQITSQPKAWIIASVAMLGFAALPGMPTGVFITIAIICGAGGLLQLQRAKPKA
                     DEQRTAAVAPEMNGKEDLRTFSPSRQFVLQFHPGQDSAQIEALVSEIRKRRNRLVVQY
                     GLTLPSFIIEHVDDIAPDEFRFTVYDVPMLKATFTQSHVAVEARQLEGENLPAAIPGN
                     TDRQEDQWVWLPAEQSGELNPVSSTTLIIERMERALQSCAPQFIGLQETKALSWLES
                     EQPELAQEMQRVLTLTRFSAVLQRLASECVPLRAIRVIAETLIEHCQHERDTNVLTDY
                     VRIALKSQIYHQYCGAEGLQVWLVTPESEGLLRDGLRQTQTETFFALSNETSQMLVQQ
                     LHIAFPVRAPEQAVLLVAQDLRSPLRTLLREEFYHVPVLSFAEISNAAKVKVMGRFDL
                     EDDLEPLDNEHAA"
BASE COUNT      772 a   1093 c   1035 g    703 t
ORIGIN


  Query Match           76.0%;  Score 19;  DB 1;  Length 3603;
  Best Local Similarity  100.0%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcat 19
           ||||||||||||||||||
Db    1522 GCAGGTGCATCGCCAGCAT 1504


RESULT  20
AF344071
LOCUS       AF344071      598 bp    DNA             STS       13-AUG-2001
DEFINITION  Oryza sativa subsp. indica microsatellite marker RM251, sequence
            tagged site.
ACCESSION   AF344071
VERSION     AF344071.1  GI:14209966
KEYWORDS    STS.
SOURCE      Oryza sativa subsp. indica.
  ORGANISM  Oryza sativa subsp. indica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 598)
  AUTHORS   Chen,X., Temnykh,S., Xu,Y., Cho,Y.G. and McCouch,S.R.
  TITLE     Development of a microsatellite framework map providing genome-wide
            coverage in rice (Oryza sativa L.)
  JOURNAL   Theor. Appl. Genet. 95, 553-567 (1997)
REFERENCE   2  (bases 1 to 598)
  AUTHORS   Temnykh,S., DeClerck,G., Lukashova,A., Lipovich,L., Cartinhour,S.
            and McCouch,S.
  TITLE     Computational and Experimental Analysis of Microsatellites in Rice
            (Oryza sativa L.): Frequency, Length Variation, Transposon
            Associations, and Genetic Marker Potential
  JOURNAL   Genome research. 11 (8), 1441-1452 (2001)
  MEDLINE   21376151
   PUBMED   11483586
REFERENCE   3  (bases 1 to 598)
  AUTHORS   Chen,X., Temnykh,S., Xu,Y., Cho,Y.G. and McCouch,S.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2001) Department of Plant Breeding, Cornell
            University, Bradfield Hall, Ithaca, NY 14853-1901, USA
FEATURES             Location/Qualifiers
     source          1. .598
                     /organism="Oryza sativa subsp. indica"
                     /cultivar="IR36"
                     /sub_species="indica"
                     /db_xref="taxon:39946"
                     /chromosome="3"
                     /clone="CT489"
                     /clone_lib="Tsp509-digested genomic"
     primer_bind     427. .444
     STS             427. .573
                     /standard_name="microsatellite marker RM251"
     primer_bind     complement(554. .573)
BASE COUNT      148 a    134 c    152 g    155 t      9 others
ORIGIN


  Query Match           75.2%;  Score 18.8;  DB 11;  Length 598;
  Best Local Similarity  90.9%;  Pred. No. 3.6e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 ggtgcatcgccagcatttcgat 25
           ||||||| ||||||||| ||||
Db     186 GGTGCATGGCCAGCATTGCGAT 207


RESULT  21
AY037869/c
LOCUS       AY037869     3078 bp    DNA             BCT       18-JUL-2001
DEFINITION  Pseudomonas chlororaphis GacS gene, complete cds.
ACCESSION   AY037869
VERSION     AY037869.1  GI:14906038
KEYWORDS    .
SOURCE      Pseudomonas chlororaphis.
  ORGANISM  Pseudomonas chlororaphis
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 3078)
  AUTHORS   Pierson,L.S. III and Chancey,S.T.
  TITLE     Pseudomonas aureofaciens 30-84 gacS
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3078)
  AUTHORS   Chancey,S.T. and Pierson,L.S. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-2001) Plant Pathology, The University of Arizona,
            Box 210036, Tucson, AZ 86721, USA
FEATURES             Location/Qualifiers
     source          1. .3078
```

```
                     /organism="Pseudomonas chlororaphis"
                     /strain="30-84"
                     /db_xref="taxon:333"
     CDS             232. .2985
                     /codon_start=1
                     /transl_table=11
                     /product="GacS"
                     /protein_id="AAK72489.1"
                     /db_xref="GI:14906039"
                     /translation="MLKKLGIKGRVLLLTLLPTSLMALVLGGYFTWMQQSDLQTQLLQ
                     RGEMIAEQLAPLVAPALAHKNIALLERIATQSLEQQDVRAVTFLAPDRTPLAHAGPSM
                     LNQAPTGNGSQMLQRTGSDATRYLLPVFGRHRNLAGDLIPDESDRLLGWVELELSHNS
                     MLLRGYRSLFASLLLIIAGLICTGLLALRMGRTINNPLSQIKHAVTQLKDGNLETRLP
                     LLGSQELDELASGINRMAGTLQNAQEELQHSIDQATEDVRQNLETIEIQNIELDLARK
                     EALEASRIKSEFLANMSHEIRTPLNGILGFTHLLQKSELTPRQLDYLGTIEKSADSLL
                     GIINEILDFSKIEAGKLVLDSIPFNLRDLLQDTLTILAPAAHAKQLELVSLVYRDTPL
                     SLVGDPLRLKQILTNLVSNAIKFTREGTIVARAMLEEEHEDSVQLRISIQDTGIGLSN
                     QDVRALFQAFSQADNSLSRQPGGTGLGLVISKRLIEQMGGEIGVDSTPGEGSEFWISL
                     RLPKTRDDAEDLPAPPLLGRRIAVLENHELARQALQHQLEDCGLEVTPFNTLEALTNG
                     VTGVHQTEQAIDLAVLGITSNDMPPERLNQHIWDLEHLGCKVLVLCPTTEQTLFHLSV
                     PNPHSQLQAKPACTRKLRRALADLVNPKVVRNEPSEPIASRPPRVLCVDDNPANLLLV
                     QTLLEDMGAKVLAVDSGYAAVKAVQNESFDLVMMDVQMPGMDGRQSTEAIRQWESERN
                     CTPLPVIALTAHAMANEKRALLQSGMDDYLTKPISERQLAQVVLKWTGLALRNQGPER
                     TGEPSGNGLELQVLDHDEGLRLAAGKADLAADMLAMLLASLEADREAIRAARAANDHN
                     ALIERVHRLHGATRYCGVPQLRAACQRSETLLKQEDVKAFAALDELERAINRLATEAR
                     INA"
BASE COUNT      643 a   1049 c    915 g    471 t
ORIGIN


  Query Match           75.2%;  Score 18.8;  DB 1;  Length 3078;
  Best Local Similarity  90.9%;  Pred. No. 3.4e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcg 23
           |||| |||||||||||||| |||
Db    2739 CAGGAGCATCGCCAGCATGTCG 2718


RESULT  22
AF192795/c
LOCUS       AF192795     3148 bp    DNA             BCT       07-NOV-1999
DEFINITION  Pseudomonas chlororaphis two component histidine sensor kinase GacS
            (gacS) gene, complete cds.
ACCESSION   AF192795
VERSION     AF192795.1  GI:6273354
KEYWORDS    .
SOURCE      Pseudomonas chlororaphis.
  ORGANISM  Pseudomonas chlororaphis
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 3148)
  AUTHORS   Kim,Y.C., Thomas,E., Kropp,B., Bauer,D. and Anderson,A.J.
  TITLE     A two-component sensor kinase (GacS) is involved in phenazine
            production of Pseudomonas chlororaphis O6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3148)
  AUTHORS   Kim,Y.C., Thomas,E., Kropp,B., Bauer,D. and Anderson,A.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Biology, Utah State University, BNR309,
            Logan, UT 84322-5305, USA
FEATURES             Location/Qualifiers
     source          1. .3148
                     /organism="Pseudomonas chlororaphis"
                     /strain="O6"
                     /db_xref="taxon:333"
     gene            256. .3009
                     /gene="gacS"
     CDS             256. .3009
                     /gene="gacS"
                     /function="phenazine production"
                     /codon_start=1
                     /transl_table=11
                     /product="two component histidine sensor kinase GacS"
                     /protein_id="AAF06332.1"
                     /db_xref="GI:6273355"
                     /translation="MLKKLGIKGRVLLLTLLPTSLMALVLGGYFTWMQQSDLQTQLLQ
                     RGEMIAEQLAPLVAPALAHQNTALLERIATQSLEQQDVRAVTFLAPDRTPLAHAGPSM
                     LNQAPTGNGSQMLQRTGSDATRYLLPVFGRHRNLAGDLIPDESDRLLGWVELELSHNS
                     MLLRGYRSLFASLMLITAGLICTGLLALRMGRTINNPLSQIKQAVTQLKDGNLETRLP
                     LLGSQELDELASGINRMAGTLQNAQEELQHSIDQATEDVRQNLETIEIQNIELDLARK
                     EALEASRIKSEFLANMSHEIRTPLNGILGFTHLLQKSELTPRQLDYLGTIEKSADSLL
                     GIINEILDFSKIEAGKLVLDSIPFNLRDLLQDTLTILAPAAHAKQLELVSLVYRDTPL
                     SLVGDPLRLKQILTNLVSNAIKFTREGTIVARAMLEEEHEDSVQLRISIQDTGIGLSN
                     QDVRALFQAFSQADNSLSRQPGGTGLGLVISKRLIEQMGGEIGVDSTPGEGSEFWISL
                     RLPKTRDDAEDLPAPPLLGRRVAVLENHELARQALQHQLEDCGLEVTPFNTLEALTNG
                     VTGVHQTEQAIDLAVLGITTNDMPPERLNQHIWDLEHLGCKVLVLCPTTEQTLFHLSV
                     PNPHSQLQAKPACTRKLRRALADLVNPKAVRSEPSEPIASRPPRVLCVDDNPANLLLV
                     QTLLEDMGAKVLAVDSGYAAVKAVQNESFDLVMMDVQMPGMDGRQSTEAIRQWESERN
                     CTPLPVIALTAHAMANEKRALLQSGMDDYLTKPISERQLAQVVLKWTGLALRNQGPER
                     TGEPSGNGLELQVLDHDEGLRLAAGKADLAADMLAMLLASLEADREAIRAARAANDHN
                     ALIERVHRLHGATRYCGVPQLRAACQRSETLLKQEDIKAFAALDELERAINRLATEAR
                     INA"
BASE COUNT      648 a   1082 c    947 g    471 t
ORIGIN


  Query Match           75.2%;  Score 18.8;  DB 1;  Length 3148;
  Best Local Similarity  90.9%;  Pred. No. 3.4e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcg 23
           |||| |||||||||||||| |||
Db    2763 CAGGAGCATCGCCAGCATGTCG 2742


RESULT  23
AC037425
LOCUS       AC037425   134058 bp    DNA             PLN       14-OCT-2000
DEFINITION  Oryza sativa chromosome 10 BAC OSJNBa0055P24 genomic sequence,
            complete sequence.
ACCESSION   AC037425
VERSION     AC037425.7  GI:10140737
KEYWORDS    HTG.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 134058)
  AUTHORS   Buell,C.R., Yuan,Q., Moffat,K.S., Hill,J.N., Jenkins,C.N.,
            Burr,P.C., Hsiao,J., Zismann,V., Pai,G., Bowman,C.L., Fujii,C.Y.,
            VanAken,S.E., Bowman,C.L., Craven,B., Utterback,T.R., Khalak,H.,
            Feldblyum,T.V., Quackenbush,J., White,O., Salzberg,S.L. and
            Fraser,C.M.
  TITLE     Oryza sativa chromosome 10 BAC OSJNBa0055P24 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 134058)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 134058)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
REFERENCE   4  (bases 1 to 134058)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
REFERENCE   5  (bases 1 to 134058)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2000) The Institute for Genomic Research, 9712
```

```
            Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
COMMENT     On Sep 15, 2000 this sequence version replaced gi:8272680.
            Address all correspondence to:rice@tigr.org

            BAC clone OSJNBa0055P24 is from Oryza sativa chromosome 10
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan and Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), and GeneSplicer
            (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org),
            searches of the complete sequence against a peptide database and
            the plant EST database at TIGR (http://www.tigr.org/tdb/tgi.shtml).
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as unknown proteins.
            Genes without protein or EST similarity, that are predicted by more
            than two gene prediction programs over most of their length are
            annotated as hypothetical proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES                Location/Qualifiers
     source             1. .134058
                        /organism="Oryza sativa"
                        /cultivar="Nipponbare"
                        /sub_species="japonica"
                        /db_xref="taxon:4530"
                        /chromosome="10"
                        /map="near S11014"
                        /clone="OSJNBa0055P24"
     repeat_region      2354. .2380
                        /rpt_family="(CAG)n"
     repeat_region      complement(2439. .2461)
                        /rpt_family="(CACTC)n"
     mRNA               complement(<2528. .>3142)
                        /gene="OSJNBa0055P24.1"
     gene               complement(2528. .3142)
                        /gene="OSJNBa0055P24.1"
                        /note="predicted by genemark.hmm"
     CDS                complement(2528. .3142)
                        /gene="OSJNBa0055P24.1"
                        /codon_start=1
                        /product="hypothetical protein"
                        /protein_id="AAG13584.1"
                        /db_xref="GI:10140752"
                        /translation="MEFVAHAAAPDSPHSDSGGGGGGMATGATSASAAGASPSRYESQ
                        KRRDWNTFGQYLRNHRPPLSLARCSGAHVLEFLRYLDQFGKTKVHAPACPFFGHPAPP
                        APCPCPLRQAWGSLDALVGRLRAAYEENGGRPENNPFGARAVRLYLREVREHQARARG
                        VSYEKKKRKKPPHPSSAAAAHDDAANGALHHHHHMPPPPPGAAA"
     repeat_region      2531. .2633
                        /rpt_family="(CGG)n"
     repeat_region      complement(3029. .3097)
                        /rpt_family="(CGG)n"
     repeat_region      complement(3235. .3360)
                        /rpt_family="(CG)n"
     repeat_region      3538. .3562
                        /rpt_family="(CGA)n"
     repeat_region      3580. .3606
                        /rpt_family="(CGG)n"
     repeat_region      complement(5977. .6187)
                        /rpt_family="putative Tourist type MITE element"
     repeat_region      6019. .6159
                        /rpt_family="Tourist_O11 MITE element from gb:U72728 Oryza
                        longistaminata receptor-like kinase protein (Xa21), family
                        member F, pseudogene sequence (1080 to 1236) 157 nt"
     repeat_region      complement(6697. .6730)
                        /rpt_family="AT_rich"
     repeat_region      complement(6865. .6921)
                        /rpt_family="(TAA)n"
     repeat_region      7806. .7955
                        /rpt_family="putative Stowaway type MITE element"
     repeat_region      complement(8868. .57532)
                        /rpt_family="Oryza sativa gene, repeat sequence Micropon-4
                        gi|4586620|dbj|AB010112.1|AB010112"
     repeat_region      complement(8874. .8909)
                        /rpt_family="(TA)n"
     repeat_region      complement(8874. .55268)
                        /rpt_family="Oryza sativa gene, repeat sequence Micropon-3
                        gi|4586621|dbj|AB010113.1|AB010113"
     repeat_region      complement(10169. .10193)
                        /rpt_family="AT_rich"
     repeat_region      complement(10345. .10429)
                        /rpt_family="Gaigin_O12 MITE element from gb:U72728 Oryza
                        longistaminata receptor-like kinase protein (Xa21), family
                        member F, pseudogene sequence (233 to 384) 152 nt"
     repeat_region      10937. .11021
                        /rpt_family="Explorer_Os3 MITE element from gb:D25363 Rice
                        genomic DNA, G1103A, sequence tagged site (66 to 225) 160
                        nt"
     repeat_region      complement(10979. .11021)
                        /rpt_family="AT_rich"
     repeat_region      12344. .47233
                        /rpt_family="Tourist-like MITE element from O.sativa DNA
                        for LRK2 gene gi|1263159|emb|X89226.1|OSLRK2GEN (1 to 540)
                        540 nt"
     repeat_region      complement(12430. .12547)
                        /rpt_family="Tourist_O11 MITE element from gb:U72728 Oryza
                        longistaminata receptor-like kinase protein (Xa21), family
                        member F, pseudogene sequence (1080 to 1236) 157 nt"
     repeat_region      12479. .12616
                        /rpt_family="putative Tourist type MITE element"
     repeat_region      complement(12803. .12824)
                        /rpt_family="AT_rich"
     repeat_region      complement(12862. .12969)
                        /rpt_family="Tourist_O11 MITE element from gb:U72728 Oryza
                        longistaminata receptor-like kinase protein (Xa21), family
                        member F, pseudogene sequence (1080 to 1236) 157 nt"
     repeat_region      12868. .13033
                        /rpt_family="putative Tourist type MITE element"
     repeat_region      complement(13218. .13316)
                        /rpt_family="AT_rich"
     repeat_region      complement(15045. .15100)
                        /rpt_family="AT_rich"
     repeat_region      complement(15858. .16053)
                        /rpt_family="Wanderer_O11 MITE element from gb:U34601
                        Oryza sativa wanderer mobile element linked to Xa21 (411
                        to 637) 227 nt"
     repeat_region      complement(15858. .16053)
                        /rpt_family="WANDERER|WANDERER Wanderer mobile element
                        linked to Xa21."
     repeat_region      complement(15861. .16038)
                        /rpt_family="Wanderer_Os2 MITE element from gb:L28995
                        Oryza sativa 3-hydroxy-3-methylglutaryl coenzyme A
                        reductase gene (696 to 925) 230 nt"
     repeat_region      complement(16306. .16507)
                        /rpt_family="WANDERER|WANDERER Wanderer mobile element
                        linked to Xa21."
     repeat_region      complement(16310. .16507)
                        /rpt_family="Wanderer_O11 MITE element from gb:U34601
                        Oryza sativa wanderer mobile element linked to Xa21 (411
                        to 637) 227 nt"
     repeat_region      complement(16337. .16481)
                        /rpt_family="Wanderer_Os1 MITE element from gb:L28995
                        Oryza sativa 3-hydroxy-3-methylglutaryl coenzyme A
                        reductase gene (4476 to 4704) 229 nt"
     repeat_region      complement(16642. .16708)
                        /rpt_family="AT_rich"
     repeat_region      complement(17251. .17298)
                        /rpt_family="AT_rich"
     mRNA               complement(join(<20332. .20394,21021. .21092,21210. .21305,
                        22519. .22567,23229. .23284,23884. .23961,24048. .24152,
                        24280. .24388,24567. .>24700))
```

```
                     /gene="OSJNBa0055P24.2"
     gene            complement(20332. .24700)
                     /gene="OSJNBa0055P24.2"
                     /note="similar to maize EST AI374558, AI795544"
     CDS             complement(join(20332. .20394,21021. .21092,21210. .21305,
                     22519. .22567,23229. .23284,23884. .23961,24048. .24152,
                     24280. .24388,24567. .24700))
                     /gene="OSJNBa0055P24.2"
                     /codon_start=1
                     /product="unknown protein"
                     /protein_id="AAG13571.1"
                     /db_xref="GI:10140739"
                     /translation="MSTAAPASATAAASLFLRRSLSSSSPRHVFLPHWVAPIVVARPA
                     RLSAVPAKRGGGGVVAAASAAAGSAQFGAGDSENPYEILGISPLDGFDQVKMAYKRRR
                     KDAESNKDAEHLFKLERAYDMVMMEQLQNRKNGVAYGSIQVSKDIKYADNQPVVPWGP
                     SLPNFLSWASHFVYGEKEYGEVEGRGIHMAKRVLRSLGLVLGSILAVSLGYTGLANFS

  Query Match             75.2%;  Score 18.8;  DB 8;  Length 134058;
  Best Local Similarity   90.9%;  Pred. No. 3.2e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 ggtgcatcgccagcatttcgat 25
           ||||||| ||||||||| ||||
Db  101294 GGTGCATGGCCAGCATTGCGAT 101315


RESULT  24
AE000198
LOCUS       AE000198    11639 bp    DNA             BCT       01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 88 of 400 of the complete
            genome.
ACCESSION   AE000198 U00096
VERSION     AE000198.1  GI:1787189
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 11639)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 11639)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 11639)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 11639)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data.  Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .11639
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     promoter        63. .89
                     /note="factor Sigma70; predicted +1 start at 1017681"
     protein_bind    86. .105
                     /note="central position to predicted promoter:1"
                     /bound_moiety="CRP predicted site"
     promoter        complement(102. .133)
                     /note="factor Sigma70; predicted +1 start at 1017680"
     gene            123. .575
                     /gene="ycbG"
                     /note="b0956"
     CDS             123. .575
                     /gene="ycbG"
                     /function="putative enzyme; Not classified"
                     /note="o150; 100 pct identical to fragment YCBG_ECOLI
                     SW:P45569 but has 111 additional N-terminal residues"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAC74042.1"
                     /db_xref="GI:1787190"
                     /translation="MKYQQLENLESGWKWKYLVKKHREGELITRYIEASAAQEAVDVL
                     LSLENEPVLVNGWIDKHMNPELVNRMKQTIRARRKRHFNAEHQHTRKKSIDLEFIVWQ
                     RLAGLAQRRGKTLSETIVQLIEDAENKEKYANKMSSLKQDLQALLGKE"
     gene            complement(651. .1691)
                     /gene="ompA"
                     /note="b0957"
     CDS             complement(651. .1691)
                     /gene="ompA"
                     /function="membrane; Outer membrane constituents"
                     /note="f346; 100 pct identical to OMPA_ECOLI SW: P02934"
                     /codon_start=1
                     /transl_table=11
                     /product="outer membrane protein 3a (II*;G;d)"
                     /protein_id="AAC74043.1"
                     /db_xref="GI:1787191"
                     /translation="MKKTAIAIAVALAGFATVAQAAPKDNTWYTGAKLGWSQYHDTGF
                     INNNGPTHENQLGAGAFGGYQVNPYVGFEMGYDWLGRMPYKGSVENGAYKAQGVQLTA
                     KLGYPITDDLDIYTRLGGMVWRADTKSNVYGKNHDTGVSPVFAGGVEYAITPEIATRL
                     EYQWTNNIGDAHTIGTRPDNGMLSLGVSYRFGQGEAAPVVAPAPAPAPEVQTKHFTLK
                     SDVLFNFNKATLKPEGQAALDQLYSQLSNLDPKDGSVVVLGYTDRIGSDAYNQGLSER
                     RAQSVVDYLISKGIPADKISARGMGESNPVTGNTCDNVKQRAALIDCLAPDRRVEIEV
                     KGIKDVVTQPQA"
     promoter        complement(1832. .1859)
                     /note="factor Sigma70; promoter ompAp2; documented +1 at
                     1019410"
     promoter        complement(1839. .1866)
                     /note="factor Sigma70; promoter ompA; documented +1 at
```

```
                1019414"
gene            complement(2048. .2557)
                /gene="sulA"
                /note="b0958"
CDS             complement(2048. .2557)
                /gene="sulA"
                /function="phenotype; Cell division"
                /note="f169; 100 pct identical to SULA_ECOLI SW: P08846"
                /codon_start=1
                /transl_table=11
                /product="suppressor of lon; inhibits cell division and
                ftsZ ring formation"
                /protein_id="AAC74044.1"
                /db_xref="GI:1787192"
                /translation="MYTSGYAHRSSSFSSAASKIARVSTENTTAGLISEVVYREDQPM
                MTQLLLLPLLQQLGQQSRWQLWLTPQQKLSREWVQASGLPLTKVMQISQLSPCHTVES
                MVRALRTGNYSVVIGWLADDLTEEEHAELVDAANEGNAMGFIMRPVSASSHATRQLSG
                LKIHSNLYH"
protein_bind    complement(2579. .2599)
                /note="central position to sulA promoter:14.5"
                /bound_moiety="LexA documented site"
promoter        complement(2594. .2622)
                /note="factor Sigma70; promoter sulA; documented +1 at
                1020168"
promoter        2685. .2713
                /note="factor Sigma70; predicted +1 start at 1020305"
gene            2776. .3405
                /gene="b0959"
CDS             2776. .3405
                /gene="b0959"
                /function="orf; Unknown"
                /note="o209; This 209 aa ORF is 31 pct identical (8 gaps)
                to 173 residues of an approx. 224 aa protein TFOX_HAEIN
                SW:P43779"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC74045.1"
                /db_xref="GI:1787193"
                /translation="MKSLSYKRIYKSQEYLATLGTIEYRSLFGSYSLTVDDTVFAMVS
                DGELYLRACEQSAQYCVKHPPVWLTYKKCGRSVTLNYYKVDESLWRNQLKLVRLSKYS
                LDAALKEKSTRNTRERLKDLPNMSFHLEAILGEVGIKDVRALRILGAKMCWLRLRQQN
                SLVTEKILFMLEGAIIGIHEAALPVARRQELAEWADSLTPKQEFPAELE"
gene            complement(3368. .5530)
                /gene="b0960"
CDS             complement(3368. .5530)
                /gene="b0960"
                /function="orf; Unknown"
                /note="f720; This 720 aa ORF is 38 pct identical (20 gaps)
                to 692 residues of an approx. 720 aa protein YHFK_HAEIN
                SW: P44289"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC74046.1"
                /db_xref="GI:1787194"
                /translation="MAFMLSPLLKRYTWNSAWLYYARIFIALCGTTAFPWWLGDVKLT
                IPLTLGMVAAALTDLDDRLAGRLRNLIITLFCFFIASASVELLFPWPWLFAIGLTLST
                SGFILLGGLGQRYATIAFGALLIAIYTMLGTSLYEHWYQQPMYLLAGAVWYNVLTLIG
                HLLFPVRPLQDNLARCYEQLARYLELKSRMFDPDIEDQSQAPLYDLALANGLLMATLN
                QTKLSLLTRLRGDRGQRGTRRTLHYYFVAQDIHERASSSHIQYQTLREHFRHSDVLFR
                FQRLMSMQGQACQQLSRCILLRQPYQHDPHFERAFTHIDAALERMRDNGAPADLLKTL
                GFLLNNLRAIDAQLATIESEQAQALPHNNDENELADDSPHGLSDIWLRLSRHFTPESA
                LFRHAVRMSLVLCFGYAIIQITGMHHGYWILLTSLFVCQPNYNATRHRLKLRIIGTLV
                GIAIGIPVLWFVPSLEGQLVLLVITGVLFFAFRNVQYAHATMFITLLVLLCFNLLGEG
                FEVALPRVIDTLIGCAIAWAAVSYIWPDWQFRNLPRMLERATEANCRYLDAILEQYHQ
                GRDNRLAYRIARRDAHNRDAELASVVSNMSSEPNVTPQIREAAFRLLCLNHTFTSYIS
                ALGAHREQLTNPEILAFLDDAVCYVDDALHHQPADEERVNEALASLKQRMQQLEPRAD
                SKEPLVVQQVGLLIALLPEIGRLQRQITQVPQETPVSA"
gene            complement(5540. .5986)
                /gene="yccF"
                /note="b0961"
CDS             complement(5540. .5986)
                /gene="yccF"
                /function="orf; Unknown"
                /note="f148; 100 pct to fragment YCCF_ECOLI SW: P37065
                (35aa) but has 113 additional C-terminal residues"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC74047.1"
                /db_xref="GI:1787195"
```

```
  Query Match             74.4%;  Score 18.6;  DB 1;  Length 11639;
  Best Local Similarity   84.0%;  Pred. No. 4.1e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||| || |||||||||| | |||
Db   10623 GCAGCTGAATCGCCAGCATGTTGAT 10647


RESULT  25
AE005286
LOCUS       AE005286    11641 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 110
            of 155.
ACCESSION   AE005286 AE005174
VERSION     AE005286.1  GI:12514140
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 11641)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 11641)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .11641
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            124. .576
                     /gene="ycbG"
                     /note="Z1306"
     CDS             124. .576
                     /gene="ycbG"
                     /function="putative enzyme; Not classified"
                     /note="Residues 1 to 150 of 150 are 100.00 pct identical
                     to residues 1 to 150 of 150 from Escherichia coli K-12
                     Strain MG1655: B0956"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAG55442.1"
                     /db_xref="GI:12514141"
```

/translation="MKYQQLENLESGWKWKYLVKKHREGELITRYIEASAAQEAVDVL
LSLENEPVLVNGWIDKHMNPELVNRMKQTIRARRKRHFNAEHQHTRKKSIDLEFIVWQ
RLAGLAQRRGKTLSETIVQLIEDAENKEKYANKMSSLKQDLQALLGKE"

gene complement(652. .1692)
/gene="ompA"
/note="Z1307"

CDS complement(652. .1692)
/gene="ompA"
/function="membrane; Cell envelop: Outer membrane constituents"
/note="Residues 1 to 346 of 346 are 100.00 pct identical to residues 1 to 346 of 346 from Escherichia coli K-12 Strain MG1655: B0957"
/codon_start=1
/transl_table=11
/product="outer membrane protein 3a (II*;G;d)"
/protein_id="AAG55443.1"
/db_xref="GI:12514142"
/translation="MKKTAIAIAVALAGFATVAQAAPKDNTWYTGAKLGWSQYHDTGF
INNNGPTHENQLGAGAFGGYQVNPYVGFEMGYDWLGRMPYKGSVENGAYKAQGVQLTA
KLGYPITDDLDIYTRLGGMVWRADTKSNVYGKNHDTGVSPVFAGGVEYAITPEIATRL
EYQWTNNIGDAHTIGTRPDNGMLSLGVSYRFGQGEAAPVVAPAPAPAPEVQTKHFTLK
SDVLFNFNKATLKPEGQAALDQLYSQLSNLDPKDGSVVVLGYTDRIGSDAYNQGLSER
RAQSVVDYLISKGIPADKISARGMGESNPVTGNTCDNVKQRAALIDCLAPDRRVEIEV
KGIKDVVTQPQA"

gene complement(2049. .2558)
/gene="sulA"
/note="Z1308"

CDS complement(2049. .2558)
/gene="sulA"
/function="phenotype; Cell division"
/note="Residues 1 to 169 of 169 are 100.00 pct identical to residues 1 to 169 of 169 from Escherichia coli K-12 Strain MG1655: B0958"
/codon_start=1
/transl_table=11
/product="suppressor of lon; inhibits cell division and ftsZ ring formation"
/protein_id="AAG55444.1"
/db_xref="GI:12514143"
/translation="MYTSGYAHRSSSFSSAASKIARVSTENTTAGLISEVVYREDQPM
MTQLLLLPLLQQLGQQSRWQLWLTPQQKLSREWVQASGLPLTKVMQISQLSPCHTVES
MVRALRTGNYSVVIGWLADDLTEEEHAELVDAANEGNAMGFIMRPVSASSHAXRQLSG
LKIHSNLYH"

gene 2777. .3406
/gene="Z1309"

CDS 2777. .3406
/gene="Z1309"
/function="orf; Unknown function"
/note="Residues 1 to 209 of 209 are 99.52 pct identical to residues 1 to 209 of 209 from Escherichia coli K-12 Strain MG1655: B0959"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55445.1"
/db_xref="GI:12514144"
/translation="MKSLSYKRIYKSQEYLATLGTIEYRSLFGSYSLTVDDTVFAMVS
DGELYLRACEQSAQYCVKYPPVWLTYKKCGRSVTLNYYRVDESLWRNQLKLVRLSKYS
LDAALKEKSTRNTRERLKDLPNMSFHLEAILGEVGIKDVRALRILGAKMCWLRLRQQN
SLVTEKILFMLEGAIIGIHEAALPVARRQELAEWADSLTPKQEFPAELE"

gene complement(3369. .5531)
/gene="Z1311"

CDS complement(3369. .5531)
/gene="Z1311"
/function="orf; Unknown function"
/note="Residues 1 to 720 of 720 are 99.86 pct identical to residues 1 to 720 of 720 from Escherichia coli K-12 Strain MG1655: B0960"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55446.1"
/db_xref="GI:12514145"
/translation="MAFMLSPLLKRYTWNSAWLYYARIFIALCGTTAFPWWLGDVKLT
IPLTLGMVAAALTDLDDRLAGRLRNLIITLFCFFIASASVELLFPWPWLFAIGLTLST
SGFILLGGLGQRYATIAFGALLIAIYTMLGTSLYEHWYQQPMYLLAGAVWYNVLTLIG
HLLFPVRPLQDNLARCYEQLARYLELKSRMFDPDIEDESQAPLYDLALANGLLMATLN
QIKLSLLTRLRGDRGQRGTRRTLHYYFVAQDIHERASSSHIQYQTLREHFRHSDVLFR
FQRLMSMQGQACQQLSRCILLRQPYQHDPHFERAFTHIDAALERMRDNGAPADLLKTL
GFLLNNLRAIDAQLATIESEQAQALPHNNDXNELADDSPHGLSDIWLRLSRHFTPESA
LFRHAVRMSLVLCFGYAIIQITGMHHGYWILLTSLFVCQPNYNATRHRLKLRIIGTLV
GIAIGIPVLWXVPSLEGQLVLLVITGVLFFAFRNVQYAHATMFITLLVLLCFNLLGEG
FEVALPRVIDTLIGCAIAWAAVSYIWPDWQFRNLPRMLERATEANCRYLDAILEQYHQ
GRDNRLAYRIARRDAHNRDAELASVVSNMSSEPNVTPQIREAAFRLLCLNHTFTSYIS
ALGAHREQLTNPEILAFLDDAVCYVDDALHHQPADEERVNEALASLKQRMQQLEPRAD
SKEPLVVQQVGLLIALLPEIGRLQRQITQVPQETPVSA"

gene complement(5541. .5987)
/gene="yccF"
/note="Z1312"

CDS complement(5541. .5987)
/gene="yccF"
/function="orf; Unknown function"
/note="Residues 1 to 148 of 148 are 99.32 pct identical to residues 1 to 148 of 148 from Escherichia coli K-12 Strain MG1655: B0961"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55447.1"
/db_xref="GI:12514146"
/translation="MRTVLNILNFVLGGFATTLGWLLATLVSIALIFTLPLTRSCWEI
TKLSLVPYGNEAIHVDELNPAGKNVLLNTGGTVLNIFWLIFFGWWLCLMHIATGIAQC
ISIIGIPVGIANFKIAAIALWPVGRRVVSVETAQAAREANARRRFE"

gene 6110. .8164
/gene="helD"
/note="Z1313"

CDS 6110. .8164
/gene="helD"
/function="enzyme; Macromolecule synthesis, modification: DNA - replication, repair, restr./modific'n"
/note="Residues 1 to 684 of 684 are 99.85 pct identical to residues 1 to 684 of 684 from Escherichia coli K-12 Strain MG1655: B0962"
/codon_start=1
/transl_table=11
/product="DNA helicase IV"
/protein_id="AAG55448.1"
/db_xref="GI:12514147"
/translation="MELKATTLGKRLAQHPYDRAVILNAGIKVSGDRHEYLIPFNQLL
AIHCKRGLVWGELEFVLPDEKVVRLHGTEWGETQRFYHHLDAHWRRWSGEMSEIASGV
LRQQLDLIATRTGENKWLTREQTSGVQQQIRQALSALPLPVNRLEEFDNCREAWRKCQ
AWLKDIESARLQHNQAYTEAMLTEYADFFRQVESSPLNPAQARAVVNGEHSLLVLAGA
GSGKTSVLVARAGWLLARGEASPEQILLLAFGRKAAEEMDERIRERLHTEDITARTFH
ALALHIIQQGSKKVPIVSKLENDTAARHELFIAEWRKQCSEKKAQAKGWRQWLTEEMQ
WSVPEGNFWDDEKLQRRLASRLDRWVSLMRMHGGAQAEMIASAPEEIRDLFSKRIKLM
APLLKAWKGALKAENAVDFSGLIHQAIVILEKGRFISPWKHILVDEFQDISPQRAALL
AALRKQNSQTTLFAVGDDWQAIYRFSGAQMSLTTAFHENFGEGERCDLDTTYRFNSRI
GEVANRFIQQNPGQLKKPLNSLTNGDKKAVTLLDESQLDALLDKLSGYAKPEERILIL
ARYHHMRPASLEKAATRWPKLQIDFMTIHASKGQQADYVLIVGLQEGSDGFPAAARES
IMEEALLPPVEDFPDAEERRLMYVALTRARHRVWALFNKENPSPFVEILKNLDVPVAR
KP"

gene complement(8196. .8663)
/gene="mgsA"
/note="Z1314"

CDS complement(8196. .8663)
/gene="mgsA"
/function="enzyme; Central intermediary metabolism: Pool, multipurpose conversions of intermed. met'm"
/note="Residues 1 to 155 of 155 are 100.00 pct identical to residues 1 to 155 of 155 from Escherichia coli K-12 Strain MG1655: B0963"
/codon_start=1
/transl_table=11
/product="methylglyoxal synthase"
/protein_id="AAG55449.1"

```
                /db_xref="GI:12514148"
                /translation="MYIMELTTRTLPARKHIALVAHDHCKQMLMSWVERHQPLLEQHV

  Query Match             74.4%;  Score 18.6;  DB 1;  Length 11641;
  Best Local Similarity   84.0%;  Pred. No. 4.1e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||| || |||||||||||  | |||
Db   10624 GCAGCTGAATCGCCAGCATGTTGAT 10648


RESULT  26
AE004044
LOCUS       AE004044    12438 bp    DNA             BCT       15-JUN-2001
DEFINITION  Xylella fastidiosa 9a5c, section 190 of 229 of the complete genome.
ACCESSION   AE004044 AE003849
VERSION     AE004044.1  GI:9107501
KEYWORDS    .
SOURCE      Xylella fastidiosa 9a5c.
  ORGANISM  Xylella fastidiosa 9a5c
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xylella.
REFERENCE   1  (bases 1 to 12438)
  AUTHORS   Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
            Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrer,H.,
            Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
            Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H.,
            Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
            Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
            Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
            Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
            and Marino,C.L.
  TITLE     The genome sequence of the plant pathogen Xylella fastidiosa. The
            Xylella fastidiosa Consortium of the Organization for Nucleotide
            Sequencing and Analysis
  JOURNAL   Nature 406 (6792), 151-157 (2000)
  MEDLINE   20365717
   PUBMED   10910347
REFERENCE   2  (bases 1 to 12438)
  AUTHORS   Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
            Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
            Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
            Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
            Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,
            Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
            Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
            Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
            Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
            Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
            Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
            Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
            Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
            Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
            Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
            Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
            Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
            Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
            de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
            Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
            Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
            Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
            Tsuhako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
            Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
            Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
            13083-970, Brazil
FEATURES             Location/Qualifiers
     source          1. .12438
                     /organism="Xylella fastidiosa 9a5c"
                     /db_xref="taxon:160492"
                     /clone="9a5c"
     gene            complement(111. .677)
                     /gene="XF2331"
     CDS             complement(111. .677)
                     /gene="XF2331"
                     /note="similar to SP|P12833 (percent identity: 43 %/query
                     alignment coverage:  87.2 %/subject alignment coverage:
                     101.2 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder/Start codon shift: -66"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydrofolate reductase type III"
                     /protein_id="AAF85130.1"
                     /db_xref="GI:9107502"
                     /translation="MDTHLKKARENNEYLQRVVRSAMIVSLIAALDRYRAIGRDNGLP
                     WHLPDDLKRFKALTLGRPVLMGRKTAHSLGRPLPGRTNLVLTRSGHVPFSEMRAVDSL
                     DVAIVQAAAEGADTLSVIGGGELYALTLPHANAMYLTHVDVAVAGADAYFPMFDAVDW
                     EIKKREPHQDDGRHAYPFTFIDYCRRLR"
     gene            complement(720. .1514)
                     /gene="XF2332"
     CDS             complement(720. .1514)
                     /gene="XF2332"
                     /note="similar to SP|O33380 (percent identity: 70 %/query
                     alignment coverage: 100.4 %/subject alignment coverage:
                     99.3 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder/Start codon shift: -84"
                     /codon_start=1
                     /transl_table=11
                     /product="thymidylate synthase"
                     /protein_id="AAF85131.1"
                     /db_xref="GI:9107503"
                     /translation="MKQYLELLNDVLVHGIQKPDRTGTGTRSVFGWQMRFDLSQGFPL
                     VTTKKLHLRSIIHELLWFLRGETNIAYLKKHQVHIWDEWADATGELGPVYGKQWRRWA
                     GADGHEIDQIRWLLEEIKRNPDSRRLVISAWNVADLPQMALVPCHALFQFYVANGKLS
                     CQLYQRSADIFLGVPFNIASYALLTHMLAQVTGLAVGDFVHTLGDAHLYANHVEQASV
                     QLERVPRPQPVLMLNPAVTDLFDFTYDDIVIEGYDPYPAIKAPVAV"
     gene            complement(1511. .2416)
                     /gene="XF2333"
     CDS             complement(1511. .2416)
                     /gene="XF2333"
                     /note="similar to SP|P37149 (percent identity: 48 %/query
                     alignment coverage:  98.0 %/subject alignment coverage:
                     101.4 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder/Start codon shift: 123"
                     /codon_start=1
                     /transl_table=11
                     /product="prolipoprotein diacylglyceryl transferase"
                     /protein_id="AAF85132.1"
                     /db_xref="GI:9107504"
                     /translation="MIYLHAIDPIAFSLGPVKVHWYGLMYLASFGAAWCLGRQRIQAG
                     RLLGVNMDGFSDLLFYAMMGVVLGGRVGYMLFYAFHDFLQEPLLLFRVWEGGMSFHGG
                     LIGVLLAVAWWSRRHRLQMFDVLDFGAPLVPVGLGFGRLGNFIGGELWGKLTHNGWGV
                     IFPRAPLSDVPAGQLAMQDVMNFVQIQEHYAAGLLGHYARHPSQLYEAFLEGLVMFIV
                     LWLFSRKPRPRYAVSGLFALLYGVFRFLVEFVRMPDNGVYVAFGWLTRGQILSLPLIV
                     IGLFLFWLSCRSPVLQPVPAPEVAK"
     gene            complement(2503. .2961)
                     /gene="XF2334"
     CDS             complement(2503. .2961)
                     /gene="XF2334"
                     /note="similar to SP|P00556 (percent identity: 54 %/query
                     alignment coverage:  73.0 %/subject alignment coverage:
                     91.0 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="diacylglycerol kinase"
                     /protein_id="AAF85133.1"
                     /db_xref="GI:9107505"
                     /translation="MAESSPFNGVIGTIFRVPVFLTSVKLMADQFGHIPRGPRRIFKA
                     AIWSLKGLRAAWLHESSFRMEVVLLLLLGPLGLWLGQTPVEKVLLVGSLFMVLAVELL
```

```
                NSAIEAVIERYGPEYHELAGRAKDMGSAAVFITLGNVVLCWVLILGPRLF"
     gene       2939. .3121
                /gene="XF2335"
     CDS        2939. .3121
                /gene="XF2335"
                /note="hypothetical protein; identified by sequence
                similarity; putative; ORF located using Glimmer/RBSfinder"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF85134.1"
                /db_xref="GI:9107506"
                /translation="MKGDDSATTTPPSSSEAIHIPMESIRNLCQKEALHNALSNTLHR
                CTDYSAQASPTSSRNI"
     gene       complement(3178. .3903)
                /gene="XF2336"
     CDS        complement(3178. .3903)
                /gene="XF2336"
                /note="similar to GI|1729783 (percent identity: 51 %/query
                alignment coverage:  90.9 %/subject alignment coverage:
                96.5 %); identified by sequence similarity; putative; ORF
                located using Glimmer/RBSfinder"
                /codon_start=1
                /transl_table=11
                /product="two-component system, regulatory protein"
                /protein_id="AAF85135.1"
                /db_xref="GI:9107507"
                /translation="MGAVMRQTKETSGLVLIVEDNRNISEMIGEYLEGRGFEVDYAVD
                GLDGYRLAVENSYDVVVLDLMLPRLDGIEVCRQLRKDARKSTPVLMLTARDTLDDKLT
                GLGFGADDYLTKPFAIQELEARLRALIRRERRQVGAEVLKLADLVLDPVSLRATRAGT
                ELQLSPIGLRLLTILMRESPRVVSRQEIEREIWGNCLPDSDTLRSHLYNLRKVVDKPF
                GRPLLHTVQSAGYRMVDIEQSMD"
     gene       3950. .4216
                /gene="XF2337"
     CDS        3950. .4216
                /gene="XF2337"
                /note="hypothetical protein; identified by sequence
                similarity; putative; ORF located using Glimmer/RBSfinder"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF85136.1"
                /db_xref="GI:9107508"
                /translation="MKILPFIGVHSVSVSQNPDTDTLNAFSNTGNNSVFITCPTNLHA
                SGRQHHTSLFPFHSKENDAPITKLIGNQAGASTATKNKPLSVQR"
     gene       complement(4371. .5492)
                /gene="XF2338"
     CDS        complement(4371. .5492)
                /gene="XF2338"
                /note="similar to GI|2622768 (percent identity: 29 %/query
                alignment coverage:  69.7 %/subject alignment coverage:
                57.9 %); identified by sequence similarity; putative; ORF
                located using Glimmer/RBSfinder"
                /codon_start=1
                /transl_table=11
                /product="chorismate mutase"
                /protein_id="AAF85137.1"
                /db_xref="GI:9107509"
                /translation="MSASPVVGIVGSAGAYGHWLGRFLRERMGLKVIGYDPADPASLS
                PAELLRCADVLVFSVPIRQTAALIGEYVAMADGTDAGQLWIDVTSIKQAPVEAMLASQ
                AEVVGLHPMTAPPKAPTLKGRVLVVCEARLHHWRPWLMQLCEALEAECVAATPEHHDR
                LMALVQAMVHACHLAQAGVLRDYLPLLGEFQSIMSFRSMSFELDMAMIARILSLNPSI
                YEDIQFGNSYVLEMLDHLLSHLLELRALVSCGDDAARTRFRADFLEANHQVFPEASLV
                AGNYNYERIGYLLADLTEPLALSVYLPEDRPGSLQRLLHVFEQHAINLSSIHSSRTPA
                GELHFRISFDPGSDHVALMRAADEIDMRRIGCVLTRPEQ"
     gene       complement(5561. .6667)
                /gene="XF2339"
     CDS        complement(5561. .6667)

Query Match             74.4%;  Score 18.6;  DB 1;  Length 12438;
Best Local Similarity   84.0%;  Pred. No. 4.1e+02;
Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||| |||||| |||| ||| ||||
Db    6540 GCAGCTGCATCACCAGGATTACGAT 6564


RESULT  27
D90734
LOCUS       D90734      13832 bp    DNA             BCT       07-FEB-1999
DEFINITION  Escherichia coli genomic DNA. (22.0 - 22.3 min).
ACCESSION   D90734 AB001340
VERSION     D90734.1  GI:1651468
KEYWORDS    Complete and shotgun sequencing; ycbG; ompA; tolG; tut; con; sulA;
            sfiA; yccF; helD; yccG; yccK; yccA; hyaA.
SOURCE      Escherichia coli(strain:K12) DNA, clone:Kohara clone #223.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
            Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
            Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
            Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
            Nishio,Y., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K.,
            Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
  TITLE     A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
  JOURNAL   DNA Res. 3 (3), 137-155 (1996)
  MEDLINE   97061202
REFERENCE   2  (sites)
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished (1996)
REFERENCE   3  (bases 1 to 13832)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1996) to the DDBJ/EMBL/GenBank databases.
            Hirotada Mori, NARA Institute of Science and Technology, Res. &
            Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01,
            Japan (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
             Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
             Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
             Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
             Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
             Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
             Yamamoto,Y. and Yano,M.
            Headed by:
             Name: Takashi Horiuchi
             Address: National Institute of Basic Biology, Okazaki, 444, Japan
             E-mail: kishori@nibb.ac.jp
            Information operator:
             Name: Hirotada Mori
             Address: NARA Institute of Science and Technology,
                      Ikoma, 630-01, Japan
             E-mail: hmori@gtc.aist-nara.ac.jp
            URL:
             The Japan E. coli genome database
             http:bsw3.aist-nara.ac.jp.
FEATURES                Location/Qualifiers
```

```
source          1. .13832
                /organism="Escherichia coli"
                /strain="K12"
                /db_xref="taxon:562"
                /clone="Kohara clone #223"
                /map="22.0-22.3 min"
                /note="Nucleotide position 1019026-1032857 from the
                initiation site of ThrA (0 min.).~This clone is from
                Kohara lambda miniset library."
gene            8. .460
                /gene="ycbG"
CDS             8. .460
                /gene="ycbG"
                /note="ORF_ID:o223#1
                similar to SwissProt Accession Number P45569"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein in ompA 3'region ."
                /protein_id="BAA35721.1"
                /db_xref="GI:4062526"
                /translation="MKYQQLENLESGWKWKYLVKKHREGELITRYIEASAAQEAVDVL
                LSLENEPVLVNGWIDKHMNPELVNRMKQTIRARRKRHFNAEHQHTRKKSIDLEFIVWQ
                RLAGLAQRRGKTLSETIVQLIEDAENKEKYANKMSSLKQDLQALLGKE"
gene            complement(536. .1576)
                /gene="con"
                /note="ompA; tolG; tut"
CDS             complement(536. .1576)
                /gene="con"
                /note="ORF_ID:o223#2
                similar to SwissProt Accession Number P02934"
                /codon_start=1
                /transl_table=11
                /product="Outer membrane protein a precursor (outer
                membrane protein II*)."
                /protein_id="BAA35722.1"
                /db_xref="GI:1651469"
                /translation="MKKTAIAIAVALAGFATVAQAAPKDNTWYTGAKLGWSQYHDTGF
                INNNGPTHENQLGAGAFGGYQVNPYVGFEMGYDWLGRMPYKGSVENGAYKAQGVQLTA
                KLGYPITDDLDIYTRLGGMVWRADTKSNVYGKNHDTGVSPVFAGGVEYAITPEIATRL
                EYQWTNNIGDAHTIGTRPDNGMLSLGVSYRFGQGEAAPVVAPAPAPAPEVQTKHFTLK
                SDVLFNFNKATLKPEGQAALDQLYSQLSNLDPKDGSVVVLGYTDRIGSDAYNQGLSER
                RAQSVVDYLISKGIPADKISARGMGESNPVTGNTCDNVKQRAALIDCLAPDRRVEIEV
                KGIKDVVTQPQA"
gene            complement(1933. .2442)
                /gene="sfiA"
                /note="sulA"
CDS             complement(1933. .2442)
                /gene="sfiA"
                /note="ORF_ID:o223#3
                similar to SwissProt Accession Number P08846"
                /codon_start=1
                /transl_table=11
                /product="Cell division inhibitor."
                /protein_id="BAA35723.1"
                /db_xref="GI:1651470"
                /translation="MYTSGYAHRSSSFSSAASKIARVSTENTTAGLISEVVYREDQPM
                MTQLLLLPLLQQLGQQSRWQLWLTPQQKLSREWVQASGLPLTKVMQISQLSPCHTVES
                MVRALRTGNYSVVIGWLADDLTEEEHAELVDAANEGNAMGFIMRPVSASSHATRQLSG
                LKIHSNLYH"
CDS             2661. .3290
                /note="ORF_ID:o223#4"
                /codon_start=1
                /transl_table=11
                /protein_id="BAA35724.1"
                /db_xref="GI:4062527"
                /translation="MKSLSYKRIYKSQEYLATLGTIEYRSLFGSYSLTVDDTVFAMVS
                DGELYLRACEQSAQYCVKHPPVWLTYKKCGRSVTLNYYRVDESLWRNQLKLVRLSKYS
                LDAALKEKSTRNTRERLKDLPNMSFHLEAILGEVGIKDVRALRILGAKMCWLRLRQQN
                SLVTEKILFMLEGAIIGIHEAALPVARRQELAEWADSLTPKQEFPAELE"
CDS             complement(3253. .5415)
                /note="ORF_ID:o223#5
                similar to PIR Accession Number B64040"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein HI1680"
                /protein_id="BAA35725.1"
                /db_xref="GI:4062528"
                /translation="MAFMLSPLLKRYTWNSAWLYYARIFIALCGTTAFPWWLGDVKLT
                IPLTLGMVAAALTDLDDRLAGRLRNLIITLFCFFIASASVELLFPWPWLFAIGLTLST
                SGFILLGGLGQRYATIAFGALLIAIYTMLGTSLYEHWYQQPMYLLAGAVWYNVLTLIG
                HLLFPVRPLQDNLARCYEQLARYLELKSRMFDPDIEDQSQAPLYDLALANGLLMATLN
                QTKLSLLTRLRGDRGQRGTRRTLHYYFVAQDIHERASSSHIQYQTLREHFRHSDVLFR
                FQRLMSMQGQACQQLSRCILLRQPYQHDPHFERAFTHIDAALERMRDNGAPADLLKTL
                GFLLNNLRAIDAQLATIESEQAQALPHNNDENELADDSPHGLSDIWLRLSRHFTPESA
                LFRHAVRMSLVLCFGYAIIQITGMHHGYWILLTSLFVCQPNYNATRHRLKLRIIGTLV
                GIAIGIPVLWFVPSLEGQLVLLVITGVLFFAFRNVQYAHATMFITLLVLLCFNLLGEG
                FEVALPRVIDTLIGCAIAWAAVSYIWPDWQFRNLPRMLERATEANCRYLDAILEQYHQ
                GRDNRLAYRIARRDAHNRDAELASVVSNMSSEPNVTPQIREAAFRLLCLNHTFTSYIS
                ALGAHREQLTNPEILAFLDDAVCYVDDALHHQPADEERVNEALASLKQRMQQLEPRAD
                SKEPLVVQQVGLLIALLPEIGRLQRQITQVPQETPVSA"
gene            complement(5425. .5871)
                /gene="yccF"
CDS             complement(5425. .5871)
                /gene="yccF"
                /note="ORF_ID:o223#6
                similar to SwissProt Accession Number P37065"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein in held 5'region ."
                /protein_id="BAA35726.1"
                /db_xref="GI:4062529"
                /translation="MRTVLNILNFVLGGFATTLGWLLATLVSIVLIFTLPLTRSCWEI
                TKLSLVPYGNEAIHVDELNPAGKNVLLNTGGTVLNIFWLIFFGWWLCLMHIATGIAQC
                ISIIGIPVGIANFKIAAIALWPVGRRVVSVETAQAAREANARRRFE"
gene            5994. .8048
                /gene="helD"
CDS             5994. .8048
                /gene="helD"
                /note="ORF_ID:o223#7
                similar to PIR Accession Number JV0021"
                /codon_start=1
                /transl_table=11
                /product="Helicase (EC 3.6.1.-) IV."
                /protein_id="BAA35727.1"
                /db_xref="GI:1651471"
                /translation="MELKATTLGKRLAQHPYDRAVILNAGIKVSGDRHEYLIPFNQLL
                AIHCKRGLVWGELEFVLPDEKVVRLHGTEWGETQRFYHHLDAHWRRWSGEMSEIASGV
                LRQQLDLIATRTGENKWLTREQTSGVQQQIRQALSALPLPVNRLEEFDNCREAWRKCQ
                AWLKDIESARLQHNQAYTEAMLTEYADFFRQVESSPLNPAQARAVVNGEHSLLVLAGA
                GSGKTSVLVARAGWLLARGEASPEQILLLAFGRKAAEEMDERIRERLHTEDITARTFH
                ALALHIIQQGSKKVPIVSKLENDTAARHELFIAEWRKQCSEKKAQAKGWRQWLTEEMQ
                WSVPEGNFWDDEKLQRRLASRLDRWVSLMRMHGGAQAEMIASAPEEIRDLFSKRIKLM
                APLLKAWKGALKAENAVDFSGLIHQAIVILEKGRFISPWKHILVDEFQDISPQRAALL
                AALRKQNSQTTLFAVGDDWQAIYRFSGAQMSLTTAFHENFGEGERCDLDTTYRFNSRI
                GEVANRFIQQNPGQLKKPLNSLTNGDKKAVTLLDESQLDALLDKLSGYAKPEERILIL
                ARYHHMRPASLEKAATRWPKLQIDFMTIHASKGQQADYVIIVGLQEGSDGFPAAARES
                IMEEALLPPVEDFPDAEERRLMYVALTRARHRVWALFNKENPSPFVEILKNLDVPVAR
                KP"
gene            complement(8080. .8547)
                /gene="yccG"
CDS             complement(8080. .8547)
                /gene="yccG"
                /note="ORF_ID:o223#8
                similar to SwissProt Accession Number P37066"
```

```
  Query Match             74.4%;  Score 18.6;  DB 1;  Length 13832;
  Best Local Similarity   84.0%;  Pred. No. 4.1e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||| || |||||||||| | |||
Db   10508 GCAGCTGAATCGCCAGCATGTTGAT 10532


RESULT  28
AC026813
```

```
LOCUS       AC026813   200631 bp    DNA             HTG       24-MAR-2000
DEFINITION  Mus musculus chromosome 5 clone RP23-119M19 strain C57BL6/J,
            WORKING DRAFT SEQUENCE, 15 unordered pieces.
ACCESSION   AC026813
VERSION     AC026813.1  GI:7321458
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 200631)
  AUTHORS   Beckstrom-Sternberg,S.M., Benjamin,B., Bouffard,G.G.,
            Dietrich,N.L., Eagle,W.O., Gan,W., Gupta,J., Ho,S.-L., Huang,M.C.,
            Idol,J., Lee-Lin,S.-Q., Maduro,Q.L., Maduro,V.B., Mastrian,S.D.,
            McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R., Stantripop,S.,
            Summers,T.J., Thomas,J.W., Thomas,P.J., Tiongson,E.E.,
            Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A., Wetherby,K.D. and
            Green,E.D.
  TITLE     NISC Mouse Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200631)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     -------------- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
            Center project name: xh
            Center clone name: 119M19
            -------------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 193127 bases at least Q40
            Consensus quality: 195226 bases at least Q30
            Consensus quality: 196033 bases at least Q20
            Insert size: 197000; agarose-fp
            Insert size: 192000; pulse-field-gel
            Insert size: 200631; sum-of-contigs
            Quality coverage: 7.36x in Q20 bases; agarose-fp
            Quality coverage: 7.55x in Q20 bases; pulse-field-gel
            Quality coverage: 7.23x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2735: contig of 2735 bp in length
            *                     gap of unknown length
            *     2736     5031: contig of 2296 bp in length
            *                     gap of unknown length
            *     5032     7390: contig of 2359 bp in length
            *                     gap of unknown length
            *     7391    10254: contig of 2864 bp in length
            *                     gap of unknown length
            *    10255    14682: contig of 4428 bp in length
            *                     gap of unknown length
            *    14683    18489: contig of 3807 bp in length
            *                     gap of unknown length
            *    18490    25669: contig of 7180 bp in length
            *                     gap of unknown length
            *    25670    38267: contig of 12598 bp in length
            *                     gap of unknown length
            *    38268    47011: contig of 8744 bp in length
            *                     gap of unknown length
            *    47012    60788: contig of 13777 bp in length
            *                     gap of unknown length
            *    60789    75761: contig of 14973 bp in length
            *                     gap of unknown length
            *    75762    89840: contig of 14079 bp in length
            *                     gap of unknown length
            *    89841   117870: contig of 28030 bp in length
            *                     gap of unknown length
            *   117871   160484: contig of 42614 bp in length
            *                     gap of unknown length
            *   160485   200631: contig of 40147 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200631
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="RP23-119M19"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    59764 a  40683 c  40100 g  60060 t     24 others
ORIGIN
```

```
  Query Match             74.4%;  Score 18.6;  DB 2;  Length 200631;
  Best Local Similarity   84.0%;  Pred. No. 3.9e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           ||||||| ||||||| ||||| ||
Db 104431 GCAGGTGGATCGCCATGATTTCAAT 104455
```

```
RESULT  29
AC068663
LOCUS       AC068663   212908 bp    DNA             HTG       01-JUN-2000
DEFINITION  Mus musculus chromosome 5 clone RP23-280D18 strain C57BL6/J,
            WORKING DRAFT SEQUENCE, 25 unordered pieces.
ACCESSION   AC068663
VERSION     AC068663.2  GI:8134859
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 212908)
  AUTHORS   Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
            Huang,M.C., Idol,J., Lee-Lin,S.-Q., Maduro,Q.L., Maduro,V.B.,
            Mastrian,S.D., McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R.,
            Stantripop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D. and Green,E.D.
  TITLE     NISC Mouse Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212908)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 1, 2000 this sequence version replaced gi:7712112.
            -------------- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
            Center project name: xi
            Center clone name: 280D18
            -------------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 199975 bases at least Q40
```

```
            Consensus quality: 202777 bases at least Q30
            Consensus quality: 204027 bases at least Q20
            Insert size: 222000; agarose-fp
            Insert size: 216000; pulse-field-gel
            Insert size: 210508; sum-of-contigs
            Quality coverage: 4.37x in Q20 bases; agarose-fp
            Quality coverage: 4.49x in Q20 bases; pulse-field-gel
            Quality coverage: 4.61x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2371: contig of 2371 bp in length
            *     2372     2471: gap of unknown length
            *     2472     4622: contig of 2151 bp in length
            *     4623     4722: gap of unknown length
            *     4723     7916: contig of 3194 bp in length
            *     7917     8016: gap of unknown length
            *     8017    10417: contig of 2401 bp in length
            *    10418    10517: gap of unknown length
            *    10518    14260: contig of 3743 bp in length
            *    14261    14360: gap of unknown length
            *    14361    18020: contig of 3660 bp in length
            *    18021    18120: gap of unknown length
            *    18121    22445: contig of 4325 bp in length
            *    22446    22545: gap of unknown length
            *    22546    26537: contig of 3992 bp in length
            *    26538    26637: gap of unknown length
            *    26638    30194: contig of 3557 bp in length
            *    30195    30294: gap of unknown length
            *    30295    34805: contig of 4511 bp in length
            *    34806    34905: gap of unknown length
            *    34906    39674: contig of 4769 bp in length
            *    39675    39774: gap of unknown length
            *    39775    44570: contig of 4796 bp in length
            *    44571    44670: gap of unknown length
            *    44671    49969: contig of 5299 bp in length
            *    49970    50069: gap of unknown length
            *    50070    55625: contig of 5556 bp in length
            *    55626    55725: gap of unknown length
            *    55726    61502: contig of 5777 bp in length
            *    61503    61602: gap of unknown length
            *    61603    70931: contig of 9329 bp in length
            *    70932    71031: gap of unknown length
            *    71032    79217: contig of 8186 bp in length
            *    79218    79317: gap of unknown length
            *    79318    88380: contig of 9063 bp in length
            *    88381    88480: gap of unknown length
            *    88481   102624: contig of 14144 bp in length
            *   102625   102724: gap of unknown length
            *   102725   112658: contig of 9934 bp in length
            *   112659   112758: gap of unknown length
            *   112759   130169: contig of 17411 bp in length
            *   130170   130269: gap of unknown length
            *   130270   149318: contig of 19049 bp in length
            *   149319   149418: gap of unknown length
            *   149419   170076: contig of 20658 bp in length
            *   170077   170176: gap of unknown length
            *   170177   189732: contig of 19556 bp in length
            *   189733   189832: gap of unknown length
            *   189833   212908: contig of 23076 bp in length.
FEATURES             Location/Qualifiers
     source          1. .212908
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="RP23-280D18"
                     /clone_lib="RPCI mouse BAC library 23"
     misc_feature    1. .2371
                     /note="assembly_fragment"
     misc_feature    2472. .4622
                     /note="assembly_fragment"
     misc_feature    4723. .7916
                     /note="assembly_fragment"
     misc_feature    8017. .10417
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    10518. .14260
                     /note="assembly_fragment"
     misc_feature    14361. .18020
                     /note="assembly_fragment"
     misc_feature    18121. .22445
                     /note="assembly_fragment"
     misc_feature    22546. .26537
                     /note="assembly_fragment"
     misc_feature    26638. .30194
                     /note="assembly_fragment"
     misc_feature    30295. .34805
                     /note="assembly_fragment"
     misc_feature    34906. .39674
                     /note="assembly_fragment"
     misc_feature    39775. .44570
                     /note="assembly_fragment"
     misc_feature    44671. .49969
                     /note="assembly_fragment"
     misc_feature    50070. .55625
                     /note="assembly_fragment"
     misc_feature    55726. .61502
                     /note="assembly_fragment"
     misc_feature    61603. .70931
                     /note="assembly_fragment"
     misc_feature    71032. .79217
                     /note="assembly_fragment"
     misc_feature    79318. .88380
                     /note="assembly_fragment"
     misc_feature    88481. .102624
                     /note="assembly_fragment"
     misc_feature    102725. .112658
                     /note="assembly_fragment"
     misc_feature    112759. .130169
                     /note="assembly_fragment"
     misc_feature    130270. .149318
                     /note="assembly_fragment"
     misc_feature    149419. .170076
                     /note="assembly_fragment"
     misc_feature    170177. .189732
                     /note="assembly_fragment"
     misc_feature    189833. .212908
                     /note="assembly_fragment"
BASE COUNT    61004 a  43094 c  43404 g  62838 t   2568 others
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 2;  Length 212908;
  Best Local Similarity   84.0%;  Pred. No. 3.9e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           ||||||| |||||||  ||||| ||
Db 145018 GCAGGTGGATCGCCATGATTTCAAT 145042


RESULT  30
AP002554
LOCUS       AP002554   327773 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 5/20.
ACCESSION   AP002554 BA000007
VERSION     AP002554.1  GI:13360491
KEYWORDS    .
```

```
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 327773)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .327773
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            150. .2258
                     /gene="ECs1032"
     CDS             150. .2258
                     /gene="ECs1032"
                     /note="similar to YCBY_ECOLI gi|1787181 percent identity
                     99 in 702 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative oxidoreductase"
                     /protein_id="BAB34455.1"
                     /db_xref="GI:13360492"
                     /translation="MNSLFASTARGLEELLKTELENLGAVECQVVQGGVHFKGDTRLV
                     YQSLMWSRLASRIMLPLGECKVYSDLDLYLGVQAINWTEMFNPGATFAVHFSGLNDTI
                     RNSQYGAMKVKDAIVDAFTRKNLPRPNVDRDAPDIRVNVWLHKETASIALDLSGDGLH
                     LRGYRDRAGIAPIKETLAAAIVMRSGWQPGTPLLDPMCGSGTLLIEAAMLATDRAPGL
                     HRGRWGFSGWAQHDEAIWQEVKAEAQTRARKGLAEYSSHFYGSDSDARVIQRARTNAR
                     LAGIGELITFEVKDVAQLTNPLPKEPYGTVLSNPPYGERLDSEPALIALHSLLGRIMK
                     NQFGGWNLSLFSASPDLLSCLQLRADKQYKAKNGPLDCVQKNYHVAESTPDSKPAMVA
                     EDYANRLRKNLKKFEKWARQEGIECYRLYDADLPEYNVAVDRYADWVVVQEYAPPKTI
                     DAHKARQRLFDIIAATISVLGIAPNKLVLKTRERQKGKNQYQKLGEKGEFLEVTEYNA
                     HLWVNLTDYLDTGLFLDHRIARRMLGQMSKGKDFLNLFSYTGSATVHAGLGGARSTTT
                     VDMSRTYLEWAERNLRLNGLTGRAHRLIQADCLAWLREANEQFDLIFIDPPTFSNSKR
                     MEDAFDVQRDHLALMKDLKRLLRAGGTIMFSNNKRGFRMDLDGLAKLGLKAQEITQKT
                     LSQDFARNRQIHNCWLITAA"
     gene            2270. .4177
                     /gene="ECs1033"
     CDS             2270. .4177
                     /gene="ECs1033"
                     /note="similar to UUP_ECOLI gi|1787182 percent identity 99
                     in 635 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative ATP-binding component of a transport
                     system"
                     /protein_id="BAB34456.1"
                     /db_xref="GI:13360493"
                     /translation="MSLISMHGAWLSFSDAPLLDNAELHIEDNERVCLVGRNGAGKST
                     LMKILNREQGLDDGRIIYEQDLIVARLQQDPPRNVEGSVYDFVAEGIEEQAEYLKRYH
                     DISRLVMNDPSEKNLNELAKVQEQLDHHNLWQLENRINEVLAQLGLDPNVALSSLSGG
                     WLRKAALGRALVSNPRVLLLDEPTNHLDIETIDWLEGFLKTFNGTIIFISHDRSFIRN
                     MATRIVDLDRGKLVTYPGNYDQYLLEKEEALRVEELQNAEFDRKLAQEEVWIRQGIKA
                     RRTRNEGRVRALKAMRRERGERREVMGTAKMQVEEASRSGKIVFEMEDVCYQVDGKQL
                     VKDFSAQVLRGDKIALIGPNGCGKTTLLKLMLGQLQADSGRIHVGTKLEVAYFDQHRA
                     ELDPDKTVMDNLAEGKQEVMVNGKPRHVLGYLQDFLFHPKRAMTPVRALSGGERNRLL
                     LARLFLKPSNLLILDEPTNDLDVETLELLEELIDSYQGTVLLVSHDRQFVDNTVTECW
                     IFEGGGKIGRYVGGYHDARGQQEQYVALKQPAVKKNEEPAAPKAETVKRSSSKLSYKL
                     QRELEQLPQLLEDLEAKLEALQTQVADASFFSQPHEQTQKVLADMAAAEQELEQAFER
                     WEYLEALKNGG"
     gene            4307. .5560
                     /gene="ECs1034"
     CDS             4307. .5560
                     /gene="ECs1034"
                     /note="similar to PQIA_ECOLI gi|1787183 percent identity
                     99 in 417 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="paraquat-inducible protein A"
                     /protein_id="BAB34457.1"
                     /db_xref="GI:13360494"
                     /translation="MCEHHHAAKHILCSQCDMLVALPRLEHGQKAACPRCGTTLTVAW
                     DAPRQRPTAYALAALFMLLLSNLFPFVNMNVAGVTSEITLLEIPGVLFSEDYASLGTF
                     FLLFVQLVPAFCLITILLLVNRAELPVRLKEQLARVLFQLKTWGMAEIFLAGVLVSFV
                     KLMAYGSIGVGSSFLPWCLFCVLQLRAFQCVDRRWLWDDIAPMPELRQPLKPGVTGIR
                     QGLRSCSCCTAILPADEPVCPRCGTKGYVRRRNSLQWTLALLVTSIMLYLPANILPIM
                     VTDLLGSKMPSTILAGVILLWSEGSYPVAAVIFLASIMVPTLKMIAIAWLCWDAKGHG
                     KRDSERMHLIYEVVEFVGRWSMIDVFVIAVLSALVRMGGLMSIYPAMGALMFALVVIM
                     TMFSAMTFDPRLSWDRQPESEHEES"
     gene            5565. .7205
                     /gene="ECs1035"
     CDS             5565. .7205
                     /gene="ECs1035"
                     /note="similar to PQIB_ECOLI gi|1787184 percent identity
                     99 in 546 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="paraquat-inducible protein B"
                     /protein_id="BAB34458.1"
                     /db_xref="GI:13360495"
                     /translation="MESNNGEAKIQKVKNWSPVWIFPIVTALIGAWVLFYHYSHQGPE
                     VTLITANAEGIEGGKTTIKSRSVDVGVVESATLADDLTHVEIKARLNSGMEKLLHKDT
                     VFWVVKPQIGREGISGLGTLLSGVYIELQPGAKGSKMDKYDLLDSPPLAPPDAKGIRV
                     VLDSKKAGQLSPGDPVLFRGYRVGSVETSTFDTQKRNISYQLFINAPYDRLVTSNVRF
```

```
                     WKDSGIAVDLTSAGMRVEMGSLTTLLSGGVSFDVPEGLDLGQPVAPKTAFVLYDDQKS
                     IQDSLYTDHIDYLMFFKDSVRGLQPGAPVEFRGIRLGTVSKVPFFAPNMRQTFNDDYR
                     IPVLIRIEPERLKMQLGENADVVEHLGELLKRGLRGSLKTGNLVTGALYVDLDFYPNT
                     PAITGIREFNGYQIIPTVSGGLAQIQQRLMEALDKINKLPLNPMIEQATSTLSESQRT
                     MKNLQTTLDSMNKILASQSMQQLPTDMQSTLRELNRSMQGFQPGSAAYNKMVADMQRL
                     DQVLRELQPVLKTLNEKSNALVFEAKDKKDPEPKRAKQ"
     gene            7217. .7765
                     /gene="ECs1036"
     CDS             7217. .7765
                     /gene="ECs1036"
                     /note="similar to YMBA_ECOLI gi|1787185 percent identity
                     98 in 182 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="hypothetical protein"
                     /protein_id="BAB34459.1"
                     /db_xref="GI:13360496"
                     /translation="MTIAALWLAGCSSGEINKNYYQLPVVQSGTQSTASQGNRLLWVE
                     QVAVPDYLAGNGVVYQTSDVKYVIANNNLWASPLDQQLRNTLVANLSTQLPGWVVASQ
                     PLGSAQDTLNVTVTEFNGRYDGKVIVSGEWLLNHQGQLIKRPFRLEGVQTQDGYDEMV
                     KVLAGVWSQEAASIAQEIKRLP"
     gene            8021. .8188
                     /gene="ECs1037"
     CDS             8021. .8188
                     /gene="ECs1037"
                     /note="identical to RMF_ECOLI gi|1787186 (Conserved in
                     E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="ribosome modulation factor"
                     /protein_id="BAB34460.1"
                     /db_xref="GI:13360497"
                     /translation="MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGW
                     REAMADRVVMA"
     gene            complement(8258. .8776)
                     /gene="ECs1038"
     CDS             complement(8258. .8776)
                     /gene="ECs1038"
                     /note="similar to FABA_ECOLI gi|1787187 percent identity
                     100 in 172 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="beta-hydroxydecanoyl thioester dehydrase"
                     /protein_id="BAB34461.1"
                     /db_xref="GI:13360498"
                     /translation="MVDKRESYTKEDLLASGRGELFGAKGPQLPAPNMLMMDRVVKMT
```

```
  Query Match             74.4%;  Score 18.6;  DB 1;  Length 327773;
  Best Local Similarity   84.0%;  Pred. No. 3.8e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||| || |||||||||||| | |||
Db   21291 GCAGCTGAATCGCCAGCATGTTGAT 21315
```

```
RESULT  31
AP002998
LOCUS       AP002998    347750 bp    DNA             BCT       15-MAY-2001
DEFINITION  Mesorhizobium loti DNA, complete genome, section 5/21.
ACCESSION   AP002998 BA000012
VERSION     AP002998.2  GI:14022369
KEYWORDS    .
SOURCE      Mesorhizobium loti (strain:MAFF303099) DNA.
  ORGANISM  Mesorhizobium loti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Phyllobacteriaceae; Mesorhizobium.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
    TITLE   Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
  JOURNAL   DNA Res. 7 (6), 331-338 (2000)
  MEDLINE   21082930
REFERENCE   2  (bases 1 to 347750)
  AUTHORS   Kaneko,T.
    TITLE   Direct Submission
  JOURNAL   Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/rhizobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
COMMENT     On May 11, 2001 this sequence version replaced gi:11994973.
FEATURES             Location/Qualifiers
     source          1. .347750
                     /organism="Mesorhizobium loti"
                     /strain="MAFF303099"
                     /db_xref="taxon:381"
     gene            complement(12. .2333)
                     /gene="mll1656"
     CDS             complement(12. .2333)
                     /gene="mll1656"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP-binding protein of dipeptide ABC
                     transporter"
                     /protein_id="BAB48979.1"
                     /db_xref="GI:14022370"
                     /translation="MNEAVRNPAEPIIEIENLSISFFTRKGEIPAVMDFSCTVMPGEA
                     MGIVGESGCGKSTVSLGIMRDLSNIGKIVGGRIKFQGKDMGELSEEELRAIRGNKIAM
                     IYQEPMASLNPAMKVGQQLMEVPLIHDKVSKEEAYKRSLEMVRAVKLPDPERMMRSYP
                     HQLSGGQQQRIVIAMALLSKPALLLLDEPTTALDVTVEAGIVELVKGLGEKFGTSMIF
                     VSHNLGLILETCDRITVMYSGEAVETGKIKDVFDRMRHPYTQGLFRSIPLPGADKNSR
                     PLISIPGQLPLPHERPKGCNFGPRCHHFVEGVCNAAEIPMVEVAGHEGHFSRCVRFNE
                     IDWEALPPGAKKVNERVVPGAPVLKIEDLRKYYKVGGSEVFGSSEGRVVKANETISFM
                     ARESETVAIVGESGCGKSTLAKVLLGLETASAGTVTLGNKQIQSTGIESRSVETVSSI
                     QMVFQNPFDTLNPSHSVGSQIIRTLEKFNVGKTVADRRKRMLELLDLVKLPRAFETRK
                     PRQLSGGQKQRIGVARAFAGDAKVVVADEPVSALDVSVQAAVTELLMDIQRKNKTTML
                     FISHDLSVVRYIADRVVVMYLGYIVEQGTTDQIFAPPYHPYTEALLSAIPIADTSVVK
                     RHIVLEGDIPSAMNPPTGCPFQTRCGYKKLVPDNLCETKVPPVKHLGDGHMSLCWLAD
                     DVLAKMEPVIKFDKEHAAHEGVPDDAPHGTGPGFAGSPPARPSGKNANAEADATGKAV
                     EDTDAVLRARRHEVRDETSETGVSKPKDTTRRH"
     gene            complement(2382. .3539)
                     /gene="mll1657"
     CDS             complement(2382. .3539)
                     /gene="mll1657"
                     /codon_start=1
                     /transl_table=11
                     /product="permease protein of dipeptide ABC transporter"
                     /protein_id="BAB48980.1"
                     /db_xref="GI:14022371"
                     /translation="MQVQYISAFTVILEVLSRFWPVWIALIIVMGASFAYKKKLALYG
                     QLFDSGVGIVGVGICLFWLFTAIFASTISPFDPLAQVPVMKDMLPGAIEPQSGLVYLF
                     GGDKLARDVFSRMVYGSQIVLIIAPAATGFALMVGITLGLPAGYYGGKIDTVLSFLAN
                     LVLAFPVILLFYLLVTPGIMDTPIPYAMAGVFFLFPIIFFSVLFWTRYKNRPDRLYIL
                     LGLTLIIGGWIYLGLVFDRDPLHIVHIDPNQLNIFVAVVFASSPGVFRIVRGLVMDIK
                     TRDYVAAAQTRGESPWYIMLWEILPNARGPLIVDACLRIGYTTILLGTLGYFGLGLAP
                     ESPDWGTAIKDASRLLRSFIHPALPPTIALMSFVLGLNLLADSLREQSMKD"
     gene            complement(3552. .4586)
                     /gene="mll1659"
     CDS             complement(3552. .4586)
                     /gene="mll1659"
                     /codon_start=1
                     /transl_table=11
                     /product="permease protein of oligopeptide ABC
                     transporter"
                     /protein_id="BAB48981.1"
                     /db_xref="GI:14022372"
                     /translation="MLSFVIRRLGTMALTMLCLTMIVFFLINLEPNLKKLAISQTEMH
```

```
                TSAEQLESWLVNHGYRQNFFVRYGQWLGVMPKQPVTDPATGKPAQRFSFCNDPVVPTF
                AGVVEGDFGCSTKFKTTVASKLFPALGATGILMFCVLAVMVPVSLLIGILAGMREGSR
                TDRVLSVASIASTATPEYVSGVIFTVIFASWLGWLNGSAASASQGITFYNFTLPVMTL
                AIYGIGYIARMTRASMVEVMTQQYIRTARLKGLSFGGVVVKHALRNALIAPFTVIMLQ
                FPWLLTGVVIVEVMFRYQGFGYTLVEAAGNNDIDLLLGCSLVSVFIVLITQLISDVGY
                AFLNPRIRVQ"
gene            complement(4809. .6464)
                /gene="mll1660"
CDS             complement(4809. .6464)
                /gene="mll1660"
                /codon_start=1
                /transl_table=11
                /product="dipeptide binding protein component of ABC
                transporter"
                /protein_id="BAB48982.1"
                /db_xref="GI:14022373"
                /translation="MTIYKSNGDRVPDHILAMAEEAKAGKVDRREFLALASVFGASTA
                MAYGLLGLAAPTEARAEDAPKKGGILKVAQWIKDPKDPRKSDWSEIANAERQALEPLV
                KYTTEYTFRPYLLESWDVNDDATQYVLHVRKGVTWHNGDAFTADDVVHNFNRWGEKAA
                EGNSMPGRLGTLFDDKTGKLREGAVTKVDDMTVKLSLTKPDISIIPGICDYVALIVHK
                SFDEKGGDFKACPIGTGPFELVSYDVGQKVVYKRRADNKWWDGDVYLDGIEFIDYGTD
                PAAMVSAFEAGEVHTNFETSADYVSILDKMDLVKSEVVTASTIVCRTNVTHKPYDNEK
                VRKALQLGVDNNVVLQLGYGNAGSVAENHHVCSIHPEYYELPKLTRDPAKAKALMAEA
                GQADFEHELITVDEDWHKNTGDAIAAQLRDAGIKVKRTVLPGSTFWNDWLKYPLSMTN
                WNMRPLGVQVLALAYRTGEAWNECAWSNPDWDAKLNAALAVSDAAKRKEMMKDIEQIL
                QDSGIIIQPYWRKLYSHSVKAVQNYKMHPTFERDYGKIWLDQA"
gene            complement(6717. .12953)
                /gene="mll1661"
CDS             complement(6717. .12953)
                /gene="mll1661"
                /note="hypothetical protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB48983.1"
                /db_xref="GI:14022374"
                /translation="MPVRCRRSRFPTSPTSRSAPASACATSRRSARCASMRPFRSTAI
                PTIRILASMPASGRRSDMKTVWRILRIVLYTLLIVVAGAIGALVVLTKTERGRDNLAG
                IISRLASSDDRKVTVSGIDGIWSGALSVDHVVLEDRQGAWLVARKVAVDWSPLALLSK
                SFSADRIAAERIELARLPVVGTQPPAQGGTTTLPVSIDIREIDLPEIALGQQLAGSGI
                AELAAKGMLKADAAPLAVETVLNVTRHDGKQGNVDAKIHFAPADDRLDLDLKASEPAG
                GIIANLLKLPDTPPVDIAVSGTGPLANWNGIATFVVDGKIVTQLTGRHQLTDKGNHVE
                AKGDGEFAPFLPDNLRSLFAGKTSFDVAGTATSAGGISIDRASIESDAIHGTATGIAD
                PAGASDLSVEIAAKSAPVQIVVGSAAQPITVAIRNATARAFGDGKAPIVDIGASLASV
                VAGGTQLNDISAQIHSDGFDIQNRAGPVAVKLTAGGLKTDVATLAPLVTGKVDVDLAG
                SLSKDAIAIDQGTLRSDALNASLTMTVALADLSMKLKMNADAVSTALPPQISSVLGER
                VKFSAAATRDPQGAFAANSISLTSGSLSASGTASAQGSDIQADIKGTLGDVSVLSPMV
                GVPVAGGVNFALSASGARTAPDFTVSADSDSLTASGRTVKAIKLTASGKADIASPSAD
                VSLTGSVNDQPLDLKAALVTSEGKRSINGFLVSLGDNKVSGDLALDDKFMPLGTLTLA
                APAIDQLAALAGQAVTGDINGTILFAKDGDAPSVTIDAKSTSIARGEVTAKAIAVNAL
                IANYLKAPAISGTVKADSVTSGATEISGIGVDLKRDGDWTNFTGGATIAGIPATAAGR
                VKIAQGTTSVEIASGEATVRGIKAAIAEPSTLTIANGTTSIEKLALDVGGGSVTLSGS
                AGQTLDLAAQFSALPAALANDFAPGLDAVGTLGGAAQVTGSPSAPEVRFNAQLAGVET
                SQTRQAELGPLAVDAAGVYTLAGGVALDHATLSGDKISGTASGTINPNGASDFALDLT
                SSGPSLPLTVGSAASPVRIEIRSLSAKVAGESTRARLDVSAILPSITTSPAKVDGLVL
                ALHSDAFDLKNRAGPVSGTVSVDKVGLDNAAIAPLIAGKVTAALNARLTSDGVAIDSG
                SLKSDALNSQVAGQVSLRDGAIDLKLSADAPSSALPAAARGMLGQRAQISATLKRDPN
                GGINVDGLKLVSGGLTAQGQASLADNKLAADITGALADVSPLSKDAKGAIAFALNAQG
                PAIAPDLSLTVNSDRLLVAEREITGLKLTATGKADAANPAANVQLTGTVAGQALQGSA
                VLATADGKSAINGLLLSLGKNRISGDLALDDKFVPEGTISLDLPDIGPLAALALEKAE
                GDVRGTIAFTKNGAAPQVAIKAATASITRGDLQAKTVSIDALIANYLAAPVISGKIRA
                DTVTSGGTVISGIDVDLKRDGDWTGFSGGATVKGIPAKAAGRVKVANGTTTIELASGQ
                ATVQGIKAAIAQASTVSIANGTTTLDRLVLNLGGGTATVTGKVGQALDINAVLAKVPM
                SLANGFSPGLDAAGSISGTVKVTGAPASPAVAFNIDAAGVQTSQTRGAGVGAVSVSSS
                GTFAGNKLTFNAN1GDGAGLGLKGGGTVTTAGTPALVLDFNGAVPFGMLSQKLAAQGL
                SLSGTANVNVQVRGPATSPVIGGSVSTSGARLVDARSGLAVNDLAADVSIGGGVARIN
                RLTGTLSTRGSLSASGTVGINPAQGFPADLSIKLTDGRYTDGRVVTANLGGDLTIKGP
                LVSAPVIAGTVNLAKTVITVPDKLPASLSALDVKHKNAPEAVRAQDKALRPATTSGKS
                GGSGLALDVTVNAPNQIFIQGRGVDAELGGSLRLTGPASSPQAVGTFTLQRGRLSILG
                KRLTFTEGTVGFSGSLVPYLNLTATTTTTSATVTIVVSGEATNPKFTFSSVPALPEDE
                VLAQLIFGRSMSNLSPLQIAQLAEAAGQLAGVGGSTSLLENLRSAIGVDDLDVTTDDK
                GGTAVSAGKYLNDRTYVTIQKGDKPGSGKATIDLNVGRGVKLRGEATDAGEAKGGIFY
                EKEY"
gene            complement(12772. .14625)
                /gene="mll1662"
CDS             complement(12772. .14625)
                /gene="mll1662"
                /note="hypothetical protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB48984.1"
                /db_xref="GI:14022375"
                /translation="MLCATALVAESRPAAAFELFGIKLWGSSNDEDADIVDPLRYAVT
                ITAPDADKDLVKRLENASALKGDEDRPVSGSLGLMAKARSDREQLVAALFADARYEGV
                VTITIDGKPLDELPPDAEFKGPQPIPVTINIAVGPKFTLGTIRLEGDAAGLMSADYGL
                ISGGDAGSGAVLKAEALIVRTLKEQGRPLAKVTDRRIVADHATSTLDVTLTVAAGPVA
                GYGATTVEGTEKVDRDFTEYMTGLKRGKQYSPQEISDARDRLLALEVFNSVTFKEADK
                LDADGNIPIGVQVSERKPRYFGLGGTFSNTEGLGLEGYWGHRNLFGHAEKLRIDGAIS
                GIGSNNLSDLNYNAGIMFEKPGVIGPASKFFAGFKTVLEHPDAYDHFSVKGSTGLSYE
                LDKQQTVSAEVALDYSRITDAFGKHTYLIASVPLQYVYDNRDSRLNPTRGFRVLAYAE
                PSYDIMSGAAFLKLKGEGSAYQSLDTASKFVLAERVAIGSIVGTGLQNVPADRRFYSG
                GGGSVRGYAYQGIGPKDFTGQPIGGLSFFETSVEMRIAVTDTIGIVPFVDAGTVSTKS
                VPNFSDVKVGAGVGLRYVTPFGPLRIDAAVPLNRDPNDPHFGIYAGIGQAF"
gene            14972. .20458
                /gene="mlr1663"
CDS             14972. .20458
                /gene="mlr1663"
                /note="hypothetical protein"
                /codon_start=1
                /transl_table=11
                /protein_id="BAB48985.1"
                /db_xref="GI:14022376"
                /translation="MAMRAARGLSILILLFFALAWNGVAQAAEARRIATTDNSDYFGF
                DLRSDQNVTLDQCKTTCLGDPACRAFTYNTKAKWCFLKSDYNQLKPFNGAVAGKVVKI
                DGDPDIGAPPELTFFPNWMADQAQQYRNRLLGPAFDKPTEGMAALISSAEQAVLTGDH
                RSAMHKYEAAVSVMPDDGQLWLNLARETLAVQPATNTSEASTLPMNATSAAFNAYKLL
                RTTKTRADVLALLGAGLDKRDLYRPALQAYEASLALVTSPAVQADYADLKARKGFRVV
                DHSVDADTSAPRICAQFSEELVKTGVDYAQFVTVDNAPPKGVEAKDKQICVEGLEHGQ
                HYDVTFRAGLPAAIGEVTAAPVVLSIYVQDRAPSARFTGDSFVLPAGARRGIPVVTVN
                MNAAEMKLYRIGDRSLAQLLSGYQFLHQLDSYDISNISEQMGAPVWQGKLDIVNDLNK
                EVTTSFPVDEALPQRKPGVYVLTAQAVDGKGDDYNSLATQWFVVSDIGLSTYTGQDGL
                NVFARSLGSAKPIAGAELTLVARNNEVLGTATSDADGHAVFNPGLTRGDGGMVPAVLM
                AKQGDNDFVFLDMSKAGFDLSDRGVTGRAAPGALDVYAWTERGIYRAGEDVHVAALAR
                DGAAKAVENLPLTFIFSRPDGVEDRRIVSDGASAGGHAVELPLEPNAMRGTWSVSIHT
                DPKQPAVASQMFLVEDFVPDRIEFDMKADKQEIERGETANINIDGRFLYGAPAAGLAL
                EGELTLSTARDWDRFPNFSFGLADEQSAEPTVTPLTNLPVVGDDGKATFPVTVDQLPS
```

Query Match 74.4%; Score 18.6; DB 1; Length 347750;
Best Local Similarity 84.0%; Pred. No. 3.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy       1 gcaggtgcatcgccagcatttcgat 25
           ||||||||| ||||| |||||||
Db 165016 GCAGGTGCACAACCAGCCTTTCGAT 165040
```

```
RESULT 32
AC013236/c
LOCUS       AC013236   119797 bp    DNA             HTG       03-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC013236
VERSION     AC013236.1  GI:6223096
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 119797)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10214372 by the submitter.
            For further information on this sequence you may e-mail to
            fly@celera.com.
```

```
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .119797
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    35319 a  25050 c  25201 g  34227 t
ORIGIN


  Query Match             73.6%;  Score 18.4;  DB 2;  Length 119797;
  Best Local Similarity   95.0%;  Pred. No. 4.9e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      6 tgcatcgccagcatttcgat 25
          ||||||||| ||||||||||
Db  17322 TGCATCGCCTGCATTTCGAT 17303


RESULT  33
AC092245/c
LOCUS       AC092245   128626 bp    DNA             INV       30-JUN-2001
DEFINITION  Drosophila melanogaster, chromosome 2L, region 35X-35X, BAC clone
            BACR32C03, complete sequence.
ACCESSION   AC092245
VERSION     AC092245.1  GI:14578126
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 128626)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2L, region 35X-35X
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 128626)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-2001) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .128626
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="35X-35X"
                     /clone="BACR32C03 (D1357)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    37802 a  27113 c  27015 g  36696 t
ORIGIN


  Query Match             73.6%;  Score 18.4;  DB 3;  Length 128626;
  Best Local Similarity   95.0%;  Pred. No. 4.8e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      6 tgcatcgccagcatttcgat 25
          ||||||||| ||||||||||
Db  72293 TGCATCGCCTGCATTTCGAT 72274


RESULT  34
AE003644
LOCUS       AE003644   264853 bp    DNA             INV       04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386055 section 37
            of 63, complete sequence.
ACCESSION   AE003644 AE002690
VERSION     AE003644.2  GI:10728786
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 264853)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
```

```
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 264853)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7298162.
FEATURES             Location/Qualifiers
     source          1. .264853
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
     mRNA            join(20111. .20584,20887. .23268)
                     /gene="noc"
                     /note="Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /product="CT14619"
                     /db_xref="FLYBASE:FBan0004491"
                     /db_xref="FLYBASE:FBgn0005771"
     gene            <20111. .>23268
                     /gene="noc"
                     /note="CG4491"
                     /map="35A4-35B1"
                     /db_xref="FLYBASE:FBan0004491"
                     /db_xref="FLYBASE:FBgn0005771"
     CDS             join(20495. .20584,20887. .22410)
                     /gene="noc"
                     /note="noc gene product; Nucleotide sequence of the Celera
                     sequence differs from the published sequence for this
                     transcript"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0004491"
                     /db_xref="FLYBASE:FBgn0005771"
                     /protein_id="AAF53399.1"
                     /db_xref="GI:7298163"
                     /translation="MVVLEGGGGVVTIGNNQYLQPDYLAPLPTTMDAKKSPLALLAQT
                     CSQIGADSSAVKPLLAMDKNKTKPGACSSSSNSSSSSGSAEISAAKSPSGQAKSPKSS
                     TPISSTATSASLSNTSTGEIKLAFKPYETNVLSHQNQNSFKSSSSLDAEPTRPSSKNS
                     SSAQERVPSRSKSNATPTDGGKAEISAHDSSSSRKTVSPSGSSQRGASPIVRSGMEVL
                     NNANGTAQHPKEMSSSMAAAAAAAAAAYKAAGPYGLNPLSALCCPPGMEQHANPAFRPP
                     FAGGFSHHHAAMLAVAANGGYPGGAPGGGPAGQPNPYISYQRIKTPAGGEAIVPVCKD
                     PYCQGCPYSAHTQQMLMGAPCPAGCTQCEHQKYGLAMASAAGLPPAHPYSQAAAAAAA
                     NAAAARSAPYVCSWVVGDAYCGKRFQTSDELFSHLRTHTGNLSDPAAAAAALAQSQAQ
                     SLLGTLFPPSALRAGYPTPPLSPMSAAAAAARYHPYAKPPPGALAGGPSPFGAAGAFN
                     PAAAAAAAALGPYYSPYAMYGQRMGAAHQ"
     mRNA            complement(join(76890. .77044,77116. .77503,77900. .>78254))
                     /gene="BG:DS04641.8"
                     /product="CT11693"
                     /db_xref="FLYBASE:FBan0004218"
                     /db_xref="FLYBASE:FBgn0028881"
     gene            complement(<76890. .>78254)
                     /gene="BG:DS04641.8"
                     /note="CG4218"
                     /map="35B1-35B2"
                     /db_xref="FLYBASE:FBan0004218"
                     /db_xref="FLYBASE:FBgn0028881"
     CDS             complement(join(76890. .77044,77116. .77503,77900. .78202))
                     /gene="BG:DS04641.8"
                     /note="BG:DS04641.8 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0004218"
                     /db_xref="FLYBASE:FBgn0028881"
                     /protein_id="AAF53400.1"
                     /db_xref="GI:7298164"
                     /translation="MATRQPTPRQLAYLDELKMRLQELRDRQASFMEQTAKILSRMSS
                     SSMDGSESETLPTDSIDENVTTNTTAGVGAGITKGSKSSLNIKKLIQRFEDLRKNSQE
                     FGDLPEVPEELINVDVRRILNGYEKLIEEGHVLQQSWFLLKKSTESCARFASANGMKS
                     EDRPPLKTNSGIVEVCNVVPEHSPEVAETPRVKKSKSLLSNMVTSSEIVDKKAEVFST
                     KMAKGKCNFQDYGHARWGAFMQLIIRVLRPFHGGSRKTKKIPAHNANNMRIPSHLPPK
                     RKPIY"
     mRNA            complement(<92373. .>92828)
                     /gene="BG:DS01486.1"
                     /product="CT11685"
                     /db_xref="FLYBASE:FBan0003473"
                     /db_xref="FLYBASE:FBgn0028913"
     gene            complement(<92373. .>92828)
                     /gene="BG:DS01486.1"
                     /note="CG3473"
                     /map="35B2-35B2"
                     /db_xref="FLYBASE:FBan0003473"
                     /db_xref="FLYBASE:FBgn0028913"
     CDS             complement(92373. .92828)
                     /gene="BG:DS01486.1"
                     /note="BG:DS01486.1 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0003473"
                     /db_xref="FLYBASE:FBgn0028913"
                     /protein_id="AAF53401.1"
                     /db_xref="GI:7298165"
                     /translation="MAALTPRIIKETQRLLEDPVPGISATPDECNARYFHVLVTGPKD
                     SPFEGGNFKLELFLPEDYPMKAPKVRFLTKIFHPNIDRVGRICLDILKDKWSPALQIR
                     TVLLSIQALLSAPNPDDPLANDVAELWKVNERRAIQLARECTLKHAMQN"
     mRNA            complement(join(129024. .130124,130936. .131047,
                     131100. .131249,131362. .131501,131576. .131710,
                     131840. .134493,139933. .140509))
                     /gene="osp"
                     /product="CT11705"
                     /db_xref="FLYBASE:FBan0003479"
                     /db_xref="FLYBASE:FBgn0003016"
     gene            complement(<129024. .>140509)
                     /gene="osp"
                     /note="CG3479"
                     /map="35B2-35B3"
                     /db_xref="FLYBASE:FBan0003479"
                     /db_xref="FLYBASE:FBgn0003016"
     CDS             complement(join(130119. .130124,130936. .131047,
                     131100. .131249,131362. .131501,131576. .131710,
                     131840. .134493,139933. .140479))
                     /gene="osp"
                     /note="osp gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0003479"
                     /db_xref="FLYBASE:FBgn0003016"
                     /protein_id="AAF53402.1"
                     /db_xref="GI:7298166"
                     /translation="MTKVLEVTSAVEVTGHPNSIAITAPERVTFVKGTSSEESQWWLN
                     ILAAFPKSKGRHKRSATLPGGQVVGSLRPTSNGDLTLSTKLGNRHSSYHKDTLTSSQS
                     AGNLLSSLDLGPSSTKTAGSPLTTTQSALADDEEDDGVETGEDVDEDEEDETSVPRKS
                     KTVSMGGQDENNRNAGNEITNRVSQPTTCLLIEDIRRDEKTIKDIANTITNLSQQQNK
                     RWSTAVNNALNNQHHFGHHSQYQVTSRDETDFQMSSNSSSTKSQNPASERPKSLPLAS
                     NSTPAIVSAIVKKIPTVMEQGDKTKPTARLQLHLKSPKHYQHERGDPDGGCNLDELCV
                     NYMAKTDELRSVGKANSKSSSGQGKPPVKEESLNAKKGWLMKQDNRTCEWSKHWFTLS
                     GAALFYYRDPLCEERGVLDGVLDVNSLTSVIPEPAASKQHAFQLTTWDKQRLVLASLS
                     PSSRNSWLAVLRSAAGLPQLDTPPKQTDIEQDFIKAQLQQPSSSPVTPGTPAGPHFSS
                     DEEYRTASEGGRRDSLDWGSPLSPSPPVLRSCLRNRSLASLHKRSRSSPPSSRRSTVD
                     SVASDELPLLVVPEEMQPTESRELKQQCETLRAEASLREARMSELLATLQRTEQQLTA
                     RLQEQQQQLNSELTQAKQSASDLMHNLGMQLTESQCQIKQLEDRLAQGIEENEGLYKR
                     LRELQAQDHSGGAALSNLQRHKIKRMDSLSDLTTISDIDPYCLQRDSLAEEYNELRSR
                     FEKAVNEIRAMKRELKQSQNQYDALELAQAALQQKLERRQHEDGAQLQLMAARIQDLT
                     LKYSSSERQVRALKQKLAKSERRRSLSLKGKEQLELKLSELQRETVERKEGTPPESSS
                     SESSSQSPLNAHLLQRLHSLEHVLLGSKERLEQSLTQLQQIRAGQRTRRSVSPMNDRK
                     DGLRQLERALAETCVMVSEQMELTCLQDSCHKCCDLRQRVEKLSALQQQTETDLQRSE
                     QLLEQRETDLAQALEKCASQEQEQELLLQQRQELSEELGRQQERCRRLEKRLELLERE
                     HGKQLECLREVYHTEHANAADEQSFRKRYQTEIEQLRTLCEKGLSAMETSHKRLTMDL
                     EQKHKMEIERLLAEKETALAEETQATLAALDAMRKAHQSEVQREVARFKQEFLRQVQR
                     GEQMRGDGAKLKEEDLGELRMEILAFSEKYSIKCVENAALEEKLHMANSKLRHFQQMQ
```

```
                     QLELRNKQFRAHLASDDPSNDVHFVQGLTSDAREDADCEDSEV"
     mRNA            join(144806. .145472,145551. .145686,145752. .146156,
                     146227. .146668)
                     /gene="Adh"
                     /note="Nucleotide sequence of the Celera sequence differs
                     from the published sequence for this transcript."
                     /product="CT11725"
                     /db_xref="FLYBASE:FBan0003481"
                     /db_xref="FLYBASE:FBgn0000055"
     gene            <144806. .>146668
                     /gene="Adh"
                     /note="CG3481"

  Query Match             73.6%;  Score 18.4;  DB 3;  Length 264853;
  Best Local Similarity   95.0%;  Pred. No. 4.8e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       6 tgcatcgccagcatttcgat 25
          ||||||||| ||||||||||
Db 100785 TGCATCGCCTGCATTTCGAT 100804


RESULT  35
DROSADH04
LOCUS       DROSADH04  300962 bp    DNA             INV       21-MAR-2000
DEFINITION  Drosophila melanogaster, chromosome 2L, region 34C4-36A7 (Adh
            region), section 4 of 10 of the complete sequence.
ACCESSION   AE003410 AC001647 AC000424 AC000421 AC000420 AC000423 AC000425
            AC000422 AC000426 AC001663 L77007 L77005 L77008 L77006 AC000602
            AC000603 AC000604 AC000605 AC000606 AC001665 L81441 L81442 L81443
            AC002440 L79841 L79849 L81431 L79843 L79845 AC000579 L79844 L79846
            AC001794 L79852 AC000580 L79839 L79848 L79851 L79854 AC000581
            AC000349 AC000582 L79842 AC000583 L81429 L79855 L81428 L79853
            AC000584 L81330 L79847 L79850 L81430 AC000585 AC000586 L79840
            AC004118 L42034 L42035 L39710 L39711 L39712 L39713 L39714 L39715
            L39716 L39717 AC000576 AC001793 AC000575 L43443 L43444 L43445
            L43446 L43447 L46823 L76993 L76994 L76995 L42036 L42037 L42038
            L39718 L39719 L39720 L39721 AC004359 L49192 L49193 L79860 L79861
            L79862 AC000591 L79863 AC000587 L76996 L76997 L48466 AC000589
            AC000588 AC000590 L43460 L43461 L43462 L43463 L43464 L43465 L43466
            L43467 L43468 L43469 L43470 L43471 L49403 L35753 L35736 L35744
            L35738 L35758 L35747 L35742 L35757 L35741 L35749 L35755 L35748
            L39673 L35761 L35759 L35735 L35740 L35746 L35751 L35737 L35752
            L35745 L35756 L35760 L39672 L35750 L35743 L35739 L35754 L49404
            L39681 L39678 L39677 L39680 L39675 L39674 L39679 L39676 AE003406
VERSION     AE003410.1  GI:7287751
KEYWORDS    HTG.
SEGMENT     4 of 10
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 300962)
  AUTHORS   Ashburner,M., Misra,S., Roote,J., Lewis,S.E., Blazej,R., Davis,T.,
            Doyle,C., Galle,R., George,R., Harris,N., Hartzell,G., Harvey,D.,
            Hong,L., Houston,K., Hoskins,R., Johnson,G., Martin,C.,
            Moshrefi,A., Palazzolo,M., Reese,M.G., Spradling,A., Tsang,G.,
            Wan,K., Whitelaw,K., Celniker,S. and Rubin,G.M.
  TITLE     An exploration of the sequence of a 2.9-Mb region of the genome of
            Drosophila melanogaster: the Adh region
  JOURNAL   Genetics 153 (1), 179-219 (1999)
  MEDLINE   99403001
   PUBMED   10471707
REFERENCE   2  (bases 1 to 300962)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2000) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, MS 64-121, Berkeley, CA
            94720, USA
COMMENT     On Mar 22, 2000 this sequence version replaced gi:2337893
            gi:1945577 gi:3006208 gi:2822114 gi:1945595 gi:2342706.
            Submitted by the Berkeley Drosophila Genome Project. For more
            information, visit the BDGP Web site: http://www.fruitfly.org/ This
            is the finished sequence of 34C4-36A7.
            The orientation of this sequence along the chromosome is left to
            right. This sequence was annotated by Sima Misra
            (sima@fruitfly.berkeley.edu) on behalf of the Berkeley Drosophila
            Genome Center. Coding sequences are predicted based on
            computational analysis, using both gene and CDS prediction programs
            and matches to other sequences. These predictions and matches have
            been evaluated by the annotators and may have been refined by hand.
             The annotators have also used their judgement about which matches
            to include in this record. The annotations on this sequence can be
            examined in more detail from
            http://www.fruitfly.org/publications/Adh.html
            The annotation syntax used in this record is documented at
            ftp://ftp.ebi.ac.uk/pub/databases/edgp/sequence_annotation.README.v
            1.2.
FEATURES                Location/Qualifiers
     source             1. .300962
                        /organism="Drosophila melanogaster"
                        /strain="y; cn bw sp"
                        /db_xref="taxon:7227"
                        /chromosome="2L"
                        /map="34C4-36A7"
                        /clone="P1 DS01368 (D100), P1 DS08249 (D25), P1 DS08284
                        (D247), P1 DS00941 (D28), P1 DS08220 (D39), P1 DS00180
                        (D29), P1 DS01514 (D33), P1 DS00131 (D35), P1 DS05899
                        (D22), P1 DS01759 (D41), P1 DS01523 (D34), P1 DS01652
                        (D43), P1 DS03792 (D31), P1 DS01068 (D37), P1 DS06238
                        (D26), P1 DS08340 (D38), P1 DS04641 (D30), P1 DS01160
                        (D42), P1 DS01486 (D32), P1 DS09219 (D101), P1 DS07721
                        (D20), P1 DS00810 (D21), P1 DS06874 (D19), P1 DS03431
                        (D102), P1 DS03144 (D106), P1 DS03323 (D127), P1 DS01219
                        (D129), P1 DS00929 (D48), P1 DS04929 (D104), P1 DS03192
                        (D103), P1 DS09194 (D267), P1 DS07295 (D47), P1 DS05639
                        (D111), P1 DS07851 (D49), P1 DS03023 (D8), P1 DS01845
                        (D7), P1 DS04862 (D107), P1 DS07108 (D51), P1 DS09217
                        (D109), P1 DS02252 (D97), P1 DS00365 (D108), P1 DS07486
                        (D46), P1 DS00913 (D24), P1 DS04095 (D112), P1 DS02795
                        (D50), P1 DS07473 (D113), P1 DS02740 (D27), P1 DS09218
                        (D114), P1 DS02780 (D23), BAC BACR48E02 (D482), BAC
                        BACR44L22 (D483)"
                        /clone_lib="P1 library, partial Sau3A in pNS582tet14Ad10
                        and RPCI-98 (Roswell Park Cancer Institute Drosophila
                        melanogaster BAC library, partial EcoRI in pBACe3.6)"
     gene               complement(11154. .14637)
                        /gene="BG:DS08340.1"
     CDS                complement(join(11154. .11215,11685. .11770,12367. .12651,
                        13069. .13190,14272. .14637))
                        /gene="BG:DS08340.1"
                        /note="symbol=BG:DS08340.1; prediction=method:''genscan'',
                        version:''1.0; Date run: 25-Jan-99; Time: 18:30:10'',
                        score:''41.800''; match=method:''BLASTN'',
                        version:''2.0a19MP-WashU [05-Feb-1998] [Build sol2.5-ultra
                        01:47:25 05-Feb-1998]'', score:''584.0'', desc:''GH12842
                        GH Drosophila melanogaster head pOT2 Drosophila
                        melanogaster cDNA clone GH12842 5prime, mRNA
                        sequence:AI135185"
                        /codon_start=1
                        /product="hypothetical protein"
                        /protein_id="AAF44876.1"
                        /db_xref="GI:7287838"
                        /db_xref="FLYBASE:FBan9998943"
                        /db_xref="FLYBASE:FBgn0004857"
                        /translation="MIRSCRALISQCGLHLRRCSLAKAASNLKKDDVEEMETPANRQR
                        VELPPNPEEKLSKRYLAFREKLRSEAPLEPLPECAPHPAHEKEPLKPWPNNTNPYTGE
```

```
                IGGQAGPEPTRYGDWERKGRKSTFQIPQIDNLANAQQAREATTSSIIHQMRFSNDLVG
                EGRWPNAYDHQSCRYEISKSHRKYLELWERVGKEAVYDQFRTHLRDFAFAESIYSWRK
                TSTDGKLEAIAGTGETTRFQSSIGRETSGKTHFTMSKETTEGVLSERITYIIDTLTLP
                VWRGGGSGGYARFFAFPAIVYHFRSRPAKG"
repeat_unit     complement(29417. .32836)
                /note="Doc element; DS08340.297"
                /function="transposon"
                /db_xref="FLYBASE:FBgn0000481"
mRNA            join(54637. .55109,55412. .57793)
                /gene="noc"
gene            54637. .57793
                /gene="noc"
CDS             join(55020. .55109,55412. .56935)
                /gene="noc"
                /note="symbol=noc; synonym=BG:DS04641.1;
                match=method:''sim4'', score:''1000.0'',
                desc:''GenBank::L14009:Drosophila melanogaster zinc finger
                protein (nocA) mRNA. CDS:385. .1998; PID:g457883.'',
                species:''Drosophila melanogaster'';
                match=method:''BLASTX'', version:''2.0a19MP-WashU
                [05-Feb-1998] [Build sol2.5-ultra 01:47:30 05-Feb-1998]'',
                score:''167.0'', desc:''trEMBL::Q90237:VITELLOGENIN
                (FRAGMENT). organism:ANOLIS PULCHELLUS. dbxref:GenBank;
                U46857; g1197667; -.'', species:''ANOLIS PULCHELLUS."
                /codon_start=1
                /evidence=experimental
                /protein_id="AAF44877.1"
                /db_xref="GI:7287839"
                /db_xref="FLYBASE:FBan9998942"
                /db_xref="FLYBASE:FBgn0005771"
                /translation="MVVLEGGGGVVTIGNNQYLQPDYLAPLPTTMDAKKSPLALLAQT
                CSQIGADSSAVKPLLAMDKNKTKPGACSSSSNSSSSSGSAEISAAKSPSGQAKSPKSS
                TPISSTATSASLSNTSTGEIKLAFKPYETNVLSHQNQNSFKSSSSLDAEPTRPSSKNS
                SSAQERVPSRSKSNATPTDGGKAEISAHDSSSSRKTVSPSGSSQRGASPIVRSGMEVL
                NNANGTAQHPKEMSSMAAAAAAAAAAAYKAAGPYGLNPLSALCCPPGMEQHANPAFRPP
                FAGGFSHHHAAMLAVAANGGYPGGAPGGGPAGQPNPYISYQRIKTPAGGEAIVPVCKD
                PYCQGCPYSAHTQQMLMGAPCPAGCTQCEHQKYGLAMASAAGLPPAHPYSQAAAAAAA
                NAAAARSAPYVCSWVVGDAYCGKRFQTSDELFSHLRTHTGNLSDPAAAAAALAQSQAQ
                SLLGTLFPPSALRAGYPTPPLSPMSAAAAAARYHPYAKPPPGALAGGPSPFGAAGAFN
                PAAAAAAAALGPYYSPYAMYGQRMGAAHQ"
tRNA            59652. .59723
                /gene="BG:DS04641.2"
                /note="symbol=BG:DS04641.2; evidence=predicted by content;
                method=tRNAscan-SE-1.0;1.0; method_score=57.22"
                /product="tRNA-Pro"
                /db_xref="FLYBASE:FBgn00"
gene            59652. .59723
                /gene="BG:DS04641.2"
tRNA            81865. .81935
                /gene="tRNA:G3:35Ba"
                /note="symbol=tRNA:G3:35Ba; synonym=BG:DS04641.3;
                evidence=predicted by content; method=tRNAscan-SE-1.0;1.0;
                method_score=71.57"
                /product="tRNA-Gly"
                /db_xref="FLYBASE:FBgn0011862"
gene            81865. .81935
                /gene="tRNA:G3:35Ba"
tRNA            complement(82081. .82151)
                /gene="tRNA:G3:35Bb"
                /note="symbol=tRNA:G3:35Bb; synonym=BG:DS04641.4;
                evidence=predicted by content; method=tRNAscan-SE-1.0;1.0;
                method_score=71.57"
                /product="tRNA-Gly"
                /db_xref="FLYBASE:FBgn0011863"
gene            complement(82081. .82151)
                /gene="tRNA:G3:35Bb"
tRNA            complement(82395. .82465)
                /gene="tRNA:G3:35Bc"
                /note="symbol=tRNA:G3:35Bc; synonym=BG:DS04641.5;
                evidence=predicted by content; method=tRNAscan-SE-1.0;1.0;
                method_score=71.57"
                /product="tRNA-Gly"
                /db_xref="FLYBASE:FBgn0011864"
gene            complement(82395. .82465)
                /gene="tRNA:G3:35Bc"
tRNA            complement(94320. .94390)
                /gene="tRNA:G3:35Bd"
                /note="symbol=tRNA:G3:35Bd; synonym=BG:DS04641.6;
```

Query Match 73.6%; Score 18.4; DB 3; Length 300962;
Best Local Similarity 95.0%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       6 tgcatcgccagcatttcgat 25
           ||||||||| ||||||||||
Db 135310 TGCATCGCCTGCATTTCGAT 135329
```

```
RESULT  36
AE004631
LOCUS       AE004631    10864 bp    DNA             BCT       30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 192 of 529 of the complete
            genome.
ACCESSION   AE004631 AE004091
VERSION     AE004631.1  GI:9948050
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 10864)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 10864)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .10864
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            complement(111. .1985)
                     /gene="PA2044"
     CDS             complement(111. .1985)
                     /gene="PA2044"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG05432.1"
                     /db_xref="GI:9948051"
                     /translation="MPRHPFSPSFLAASVFCAATPALAALQPIAEAPLAGETEFRCQF
                     NADNSTDCVSTYSFTILKPSGREMLSRIDRSYAETDSLIVEKAELTQPGGKPVPLDQS
                     QIDTRTAPNPDQGFLRERQTSLAFPNLRVGTRISYTLREHFTAKPLSTQFHYILSRPP
                     MPVRDDRFVAEFKAERPIFVRSELMDAYRIEQSADKKTLKVSLKKPQYTNYINEAGNA
                     YLRHTPRLELGSSLDLQDNFGPFAARYNEILAAELPKGAAAAVAAVKGKPAREQVAGL
                     MQYINDTYRYLGDWRASERGYVPFSLAEIERNGYGDCKDLAILLAAMLKAAGIKAEPT
                     LVSRGDVVWDLLVPGMYAPNHAIVRAEVDGKTWWLDPTNPVFAPGRIMPDIQQRWALV
                     LGADGKVRRDEIPLEAPGDTLRVTRSEHYTHDGEARVESRVELSNAPLMQLSVADRQR
                     GRTSTDQDLCRNFAKEGSDCVLERDDSQFVLPPSYTISARLTDRRALDRLGGEYFYNR
                     QDLASQWDAFAKYRSEGQLAELYLGEPQITSYDISLSGGKTDEPAHSCEIRSPWFDID
```

```
                LQAEPAKDGGYHYRYREVQKTSWLNHDEINSAEFGKLIEQSRGCVEQLRLVVKLDKRP
                "
     gene       complement(2160. .2420)
                /gene="PA2045"
     CDS        complement(2160. .2420)
                /gene="PA2045"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAG05433.1"
                /db_xref="GI:9948052"
                /translation="MKFLLIGLIRFYQYAISPLIGPRCRFYPSCSHYTLEAIRVHGAL
                RGGYLGARRLLRCHPWHPGGYDPVPERQEQACACHRTAKPGK"
     gene       complement(2479. .2886)
                /gene="PA2046"
     CDS        complement(2479. .2886)
                /gene="PA2046"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG05434.1"
                /db_xref="GI:9948053"
                /translation="MGATRQRLVELQEAFADPRRPLLVIPQPRRTIGFATAASEVRGG
                FYVLADGPFRQAYLGARWNVGELTLDGVDYLEGDAEAYCPGYGRSALIPLRGGAWREG
                TLAPEEPGLRGRLATARSLVDGRILYCAAENAP"
     gene       complement(3921. .4910)
                /gene="PA2047"
     CDS        complement(3921. .4910)
                /gene="PA2047"
                /codon_start=1
                /transl_table=11
                /product="probable transcriptional regulator"
                /protein_id="AAG05435.1"
                /db_xref="GI:9948054"
                /translation="MPVAATRASLGKVSPASPRMSENTPLLSVDLPAELRQRQSELAQ
                SLLRHCPRDGIYGTAVEPLALIRADGPTLPARGMHKPALCIIAQGRKEVQLAEERYIY
                DPLHYLVVSVTLPLAGQVIDASPDAPYLCVRLDIDPAEITQLISDAGPMEVASRRGDR
                GLYVDRIDASLLDAVLRLIHLLDSPRDIPMLAPLILREIFYRLLRSGQGQRLHEIAIA
                DSQAHRITRAIDWINQNYGKPLRIEQLAQVVNLSPSTLHHRFKAVTAMSPLQYQKQLR
                LQEARRLIFSEGLEVAAAGYRVGYESPSQFSREYSRLFGAPPLRDLARARNTA"
     gene       5324. .5731
                /gene="PA2048"
     CDS        5324. .5731
                /gene="PA2048"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG05436.1"
                /db_xref="GI:9948055"
                /translation="MRHPYTPQRPAGVSAPPALPLLVPLAQRRIHVEEAESMPAFNRA
                SHLVSIRARSGQSHRLGLRLQELAQAGQAAPGCLRYELRQADGDADLWLLHSEWSDEA
                AMQAYLSGDAQRVFAEVLWQALAATLDVQERPF"
     gene       complement(5789. .7222)
                /gene="PA2049"
     CDS        complement(5789. .7222)
                /gene="PA2049"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG05437.1"
                /db_xref="GI:9948056"
                /translation="MLVAIHGLIHRLGLYGEHLEGALVNDSEALYAHPATYYRREDLL
                PAFESNEALRERFFGGVKKPIFTSANAHNHLVGLAEVLAMYLLVWSPAPWRLVRLEPP
                PGLDARHAQRFAGERRHIEAFRDGLPRVARALAHISSLMIFDDHDITDDWNLSARWEL
                TAYGHPFSRRIIGNALLAYLLCQGWGNDPQHCGEWLAPVAALMEGAAADGRLPCAEQD
                RLLDALLDFQHWHYRLPGSPTLIVLDTRTRRWRSERNLGRPSGLMDWEALSEFQQDLL
                GEKAAIVSPAPMFGVKLIEGVQKLFSWIGQPLLVDAENWMAHRGAASVMMNIFRHSR
                TPGNYLILSGDVHYSFVYDIRIRHRERGPRLWQITSSGIKNEFPRRLLDWFDRLNRWL
                YAPWSPLNWLTQRRSLEIRPRVPQHSQAGERLWNSAGIGQVFIGEDGRPRAVLQVNAD
                GSPATAFERHRDVEDERADQSPVAPGR"
     gene       8111. .8617
                /gene="PA2050"
     CDS        8111. .8617
                /gene="PA2050"
                /codon_start=1
                /transl_table=11
                /product="probable sigma-70 factor, ECF subfamily"
                /protein_id="AAG05438.1"
                /db_xref="GI:9948057"
                /translation="MGMDQTRNRLVGLMFQNDYSWLSGRLRRYLGCPHSAEDIASETF
                LKVLALPDPASIREPRALLTTIARRLMYDGWRRQDLERAYLESLAGLPEALAPSAEEQ
                AQVVETLLRLERMLAGLSPKARAAFIHSQIGGLTYVEIACLLEVSVSRIHQYMVEGFK
                QCYQVLAE"
     gene       8614. .9567
                /gene="PA2051"
     CDS        8614. .9567
                /gene="PA2051"
                /codon_start=1
                /transl_table=11
                /product="probable transmembrane sensor"
                /protein_id="AAG05439.1"
                /db_xref="GI:9948058"
                /translation="MSQGRRERILDEAAQWMALLQSGHAGAEERLAFQQWRQADAEHG
                AVFDRLSLGLQSFRNEPLHGLRAEQLIGALEAPSARRRFLQRSLALAAFGVATLALGR
                IGIVGFPWPGDLYTGTGQRQSFRLEDGSQLVLDARSRVRPEFDGYRRRLRLREGRLLV
                DVARDPARPFVVDTGRGQVEAQGTRFQVSDDGASSRVVVLESRVRITTADGRQCLLER
                GQSARFDASAILARGPAEGGESAWLDGRLEVRDRPLGEVLEALRAYRRGIISVADDAA
                ALRVSGIYPLDDSGVALALLEQSLPIRLDYHGPYWVSIASR"
     gene       complement(9605. .10075)
                /gene="cynS"
                /note="PA2052"
     CDS        complement(9605. .10075)
                /gene="cynS"
                /codon_start=1
                /transl_table=11
                /product="cyanate lyase"
                /protein_id="AAG05440.1"
                /db_xref="GI:9948059"
                /translation="MQHSQVSPNARQQLAETVVLNKARLGLSWQDLADGTGLALTFVT
                AALLGQHALPEAAARKVAAQLGLDDDAVLLLQSIPLRGSIPGGIPSDPTIYRFYEMLQ
                VYGSTLKALVHEQFGDGIISAINFKLDIKKVEDPEGGSRAVITLDGKYLPTKPF"
     gene       complement(10115. .10777)
                /gene="cynT"
                /note="PA2053"
     CDS        complement(10115. .10777)
                /gene="cynT"
                /codon_start=1
                /transl_table=11
                /product="carbonate dehydratase"
                /protein_id="AAG05441.1"
                /db_xref="GI:9948060"
                /translation="MRDIIDGFLRFQRDAYPARSQLFKSLATRQAPKALFIACSDSRV
                VPELLTQREPGELFVIRNAGNIVPGYGPQPGGVSASVEYAVAVLGVGDIVVCGHSDCG
                AMGAIASCACLDQLPAVAGWLHHAEAARAMNSAHEHASEAARLDALVRHNVIAQLANL
                RTHPCVARALEQGRLNLHGWVYDIESGRIDALDGASRRFVSLAEHPEVRAVGGEPGQA
                VA"
BASE COUNT     1603 a   3503 c   3970 g   1788 t
ORIGIN


  Query Match           72.8%;  Score 18.2;  DB 1;  Length 10864;
  Best Local Similarity  87.0%;  Pred. No. 6.4e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcg 23
          |||||| ||||||||||| |||
Db   6964 GCAGGTACATCGCCAGCACCTCG 6986


RESULT  37
AE007132/c
LOCUS       AE007132    14493 bp    DNA             BCT       27-APR-2001
DEFINITION  Mycobacterium tuberculosis CDC1551,  section 218 of 280 of the
            complete genome.
ACCESSION   AE007132 AE000516
VERSION     AE007132.1  GI:13882943
```

```
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis CDC1551.
  ORGANISM  Mycobacterium tuberculosis CDC1551
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 14493)
  AUTHORS   Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
            Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
            Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
            Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
            Gill,J., Mikula,A. and Bishai,W.
  TITLE     Whole genome comparison of Mycobacterium tuberculosis clinical and
            laboratory strains
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 14493)
  AUTHORS   Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
            Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
            Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
            Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
            Gill,J., Mikula,A. and Bishai,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .14493
                     /organism="Mycobacterium tuberculosis CDC1551"
                     /strain="CDC1551"
                     /db_xref="taxon:83331"
                     /note="clinical strain"
     gene            complement(381. .1913)
                     /gene="MT3146"
     CDS             complement(381. .1913)
                     /gene="MT3146"
                     /note="similar to PID:455922 GB:U00096 SP:P06957
                     PID:1786303 PID:434009; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, GntR family"
                     /protein_id="AAK47479.1"
                     /db_xref="GI:13882944"
                     /translation="MGDSASARALDSANTVHSIEMSTEPDAVWTDKRASKIARRIEAD
                     IVRRGWPIGASLGSESALQQRFCVSRSVLREAVRLVEHHQVARMRRGPNGGLFICEPN
                     AGPATRAVVIYLEYLGTTIGDLLGARLVLEPLAASLAAEHIDEPGIERLRAVLRAEER
                     WRPGLPPPPEQFYRVLAEQSKNPVLQLFIDILMRLTKRYVQKSGTQSAGEAVEAAGQV
                     HNEHSDIVAAVTAGDSAWAKTLSERHVEAVAGWLQQHQRGNDAAVRNGGRAREPRRAQ
                     QLILGAPRGKLAEVLAATIGDDIAASGWQVGSVFGTETALLERYQVSRAVLREAVRLL
                     EYHAIAHMRRGPGGGLVVTTPQPQASIDTIALYLQYRKPSREDLRCVRDAIEIDNVAK
                     VVKRRSEPEVASFLDTLGRPRLDNPTDDVRAAAVEEFRFHVGLARAAGNTMLDLFLLI
                     LVELFRRHLSSTEQALPTWSDVVAVGHAHVRILEAIGSGDDSLARCRTRRHLDAAASW
                     WL"
     gene            complement(1970. .4135)
                     /gene="MT3147"
     CDS             complement(1970. .4135)
                     /gene="MT3147"
                     /note="identified by match to PFAM  protein family HMM
                     PF00441"
                     /codon_start=1
                     /transl_table=11
                     /product="acyl-CoA dehydrogenase, putative"
                     /protein_id="AAK47480.1"
                     /db_xref="GI:13882945"
                     /translation="MGIALTDDHRELSGVARAFLTSQKVRWAARASLDAAGDARPPFW
                     QNLAELGWLGLHIDERHGGSGYGLSELVVVIEELGRAVAPGLFVPTVIASAVVAKEGT
                     DDQRARLLPALIDGTLTAGVGLDSQVQVTDGVADGEAGIVLGAGLAELLLVAAGDDVL
                     VLERGRKGVSVDVPENFDPTRRSGRVRLDNVRVTTDDILLGAYESALARARTLLAAEA
                     VGGAADCVDSAVAYAKVRQQFGRTIATFQAVKHHCANMLVAAESAIAAVWDAARAAAE
                     DEEQFRLAAAVAAALAFPAYARNAELNIQVHGGIGFTWEHDAHLHLRRALVTVGLFGG
                     DAPVRDVFERTAAGVTRAISLDLPAQAEELRARIRSDAAEIAALEKDAQRDKLIETGY
                     VMPHWPRPWGRAAGAVEQLVIEEEFSAAGIERPDYSITGWVILTLIQHGTPWQIERFV
                     EKALRQQEIWCQLFSEPDAGSDAASVKTRATRVEGGWKINGQKVWTSGAQYCARGLAT
                     VRTDPDAPKHAGITTVIIDMLAPGVEVRPLRQITGDSEFNEVFFNDVFVPDEDVVGAP
                     NSGWTVARATLGNERVSIGGSGSYYEAMAAKLVQLVQRRSDAFAGAPIRVGAFLAEDH
                     ALRLLNLRRAARSVEGAGPGPEGNITKLKVAEHMIEGAAIAAALWGPEIALLDGPGRV
                     IGRTVMGARGMAIAGGTSEVTRNQIAERILGMPRDPLIS"
     gene            4292. .5815
                     /gene="MT3148"
     CDS             4292. .5815
                     /gene="MT3148"
                     /note="similar to GB:Pyro_h; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="DNA ligase"
                     /protein_id="AAK47481.1"
                     /db_xref="GI:13882946"
                     /translation="MLLHDVAITSMDVAATSSRLTKVARIAALLHRAAPDTQLVTIIV
                     SWLSGELPQRHIGVGWAALRSLPPPAPQPALTVTGVDATLSKIGTLSGKGSQAQRAAL
                     VAELFSAATEAEQTFLLRLLGGELRQGAKGGIMADAVAQAAGLPAATVQRAAMLGGDL
                     AAAAAAGLSGAALDTFTLRVGRPIGPMLAQTATSVHDALERHGGTTIFEAKLDGARVQ
                     IHRANDQVRIYTRSLDDVTARLPEVVEATLALPVRDLVADGEAIALCPDNRPQRFQVT
                     ASRFGRSVDVAAARATQPLSVFFFDILHRDGTDLLEAPTTERLAALDALVPARHRVDR
                     LITSDPTDAANFLDATLAAGHEGVMAKAPAARYLAGRRGAGWLKVKPVHTLDLVVLAV
                     EWGSGRRRGKLSNIHLGARDPATGGFVMVGKTFKGMTDAMLDWQTTRFHEIAVGPTDG
                     YVVQLRPEQVVEVALDGVQRSSRYPGGLALRFARVVRYRADKDPAEADTIDAVRALY"
     gene            5951. .8227
                     /gene="MT3149"
     CDS             5951. .8227
                     /gene="MT3149"
                     /note="similar to SP:P15078 GB:X52904 PID:41081 GB:U00096
                     PID:1651245; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="carbon starvation protein A"
                     /protein_id="AAK47482.1"
                     /db_xref="GI:13882947"
                     /translation="MAAPTPSNRIEERSGHASCVRADADLPPVAILGRSPITLRHKIF
                     FVAVAVIGALAWTVVAFFRNEPVNAVWIVVAAGCTYIIGFRFYARLIEMKVVRPRDDH
                     ATPAEILDDGTDYVPTDRRVVFGHHFAAIAGAGPLVGPVLATQMGYLPSSIWIVVGAV
                     LAGCVQDYLVLWISVRRRGRSLGQMVRDELGATAGVAALVGIPVIITIVIAVLALVVV
                     RALAKSPWGVFSIAMTIPIAIFMGCYLRFLRPGRVSEVSLIGIGLLLLAVVSGDWVAH
                     TSWGAAWFSLSPVTLCWLLISYGFAASVLPVWLLLAPRDYLSTFMKVGTIALLAIGVC
                     AAHPIIEAPAVSKFAGSGNGPVFAGSLFPFLFITIACGALSGFDALICSGTTPKMLEK
                     EGQMRVIGYGGMMTESFVAVIALLTAAILDQHLYFTLNAPSLHTHDSAATAAKYVNGL
                     GLTGSPVTPDHISQAAASVGEQTIVSRTGGAPTLAFGMAEMLHRVVGGVGLKAFWYHF
                     AIMFEALFILTTVDAGTRAARFMISDALGNFGGVLRKLQNPSWRPGAWACSLVVVAAW
                     GSILLLGVTDPLGGINTLFPLFGIANQLLAGIALTVITVVVIKKGRLKWAWIPGIPLL
                     WDLAVTLTASWQKIFSADPSVGYWTQHAHYAAAQHAGETAFGSATNADEINDVVRNTF
                     VQGTLSIVFVVVVVLVVVAGVIVALKTIRGRGIPLAEDDPAPSTLFAPAGLIPTAAER
                     KLQRRLGAPASASVAAPD"
     gene            8276. .8458
                     /gene="MT3149.1"
     CDS             8276. .8458
                     /gene="MT3149.1"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK47483.1"
                     /db_xref="GI:13882948"
                     /translation="MGPTRDSRQPAERATKRRYRFARGLMAGLLGTMALTSGGGVARE
                     DPLEPDPLAPIIDDSR"
     gene            complement(8533. .8958)
                     /gene="MT3150"
     CDS             complement(8533. .8958)
                     /gene="MT3150"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK47484.1"
                     /db_xref="GI:13882949"
                     /translation="MVKDLDRRLAGCLPAVLSLFRLVYGLLFAGYGSMILFGWPVTSA
                     QPVEFGSWPGWYAGVIELVAGLLIATGLFTRAVAFVASGEMAVAYFWMHQPYALWPIG
                     GPPDGNGGTPAILFCFGFFLLVFTGGGIYSIDARRTVTA"
```

```
     gene            9095. .9418
                     /gene="MT3150.1"
     CDS             9095. .9418
                     /gene="MT3150.1"
                     /note="similar to SP:P23895; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="emrE protein"
                     /protein_id="AAK47485.1"
                     /db_xref="GI:13882950"
                     /translation="MIYLYLLCAIFAEVVATSLLKSTEGFTRLWPTVGCLVGYGIAFA
                     LLALSISHGMQTDVAYALWSAIGTAA1VLVAVLFLGSPISVMKVVGVGLIVVGVVTLN
                     LAGAH"
     gene            9415. .10023
                     /gene="MT3151"
     CDS             9415. .10023
                     /gene="MT3151"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK47486.1"
                     /db_xref="GI:13882951"
                     /translation="MTAGSDRRPRDPAGRRQAIVEAAERVIARQGLGGLSHRRVAAEA
                     NVPVGSTTYYFNDLDALREAALAHAANASADLLAQWRSDLDKDRDLAATLARLTTVYL
                     ADQDRYRTLNELYMAAAHRPELQRLARLWPDGLLALLEPRIGRRAANAVTVFFDGATL
                     HALITGTPLSTDELTDAIARLVADGPEQREVGQSAHAGRTPD"
     gene            10136. .10546
                     /gene="MT3152"
     CDS             10136. .10546
                     /gene="MT3152"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK47487.1"
                     /db_xref="GI:13882952"
                     /translation="MLTVGVGIGAAILLGWFTLAHRHPDQPGAAATPPPAGLTTRSAP
                     TAAPPSTLQSPDLDSVFLGNLHDRGISFTNPDAAVYNGKMVCTNLGGGMTVQQVVEAL
                     QSSSPALGDRTTAYVAVSIRTYCPKYDAVLPPGS"
     tRNA            complement(10548. .10620)
                     /gene="tRNA-Ala-3"
                     /product="tRNA-Ala"
     gene            complement(10548. .10620)
                     /gene="tRNA-Ala-3"
     gene            complement(10687. .12330)
                     /gene="MT3153"
     CDS             complement(10687. .12330)

  Query Match             72.8%;  Score 18.2;  DB 1;  Length 14493;
  Best Local Similarity   87.0%;  Pred. No. 6.3e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 caggtgcatcgccagcatttcga 24
          |||||||| |||||||| | |||
Db  12759 CAGGTGCACCGCCAGCAATCCGA 12737


RESULT  38
D85415/c
LOCUS       D85415      25404 bp    DNA             BCT       13-JUL-2001
DEFINITION  Pseudomonas putida UCC22 plasmid pTDN1 genes for conversion of
            aniline to catechol, aniline degradation lower-pathway gene
            cluster, regulatory protein and transposase, partial and complete
            cds.
ACCESSION   D85415
VERSION     D85415.2  GI:14715441
KEYWORDS    .
SOURCE      Pseudomonas putida (strain:UCC22) plasmid:pTDN1 DNA.
  ORGANISM  Pseudomonas putida
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 9239)
  AUTHORS   Fukumori,F. and Saint,C.P.
  TITLE     Nucleotide sequences and regulational analysis of genes involved in
            conversion of aniline to catechol in Pseudomonas putida
            UCC22(pTDN1)
  JOURNAL   J. Bacteriol. 179 (2), 399-408 (1997)
  MEDLINE   97144524
REFERENCE   2  (bases 1 to 9239)
  AUTHORS   Fukumori,F. and Saint,C.P.
  TITLE     Aniline degradation in Pseudomonas putida UCC22(pTDN1): Initial
            characterization of its conversion to catechol
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 25404)
  AUTHORS   Fukumori,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-1996) Fumiyasu Fukumori, Toyo University, Faculty
            of Life Sciences; 1-1-1 Izumino, Itakura, Ora, Gunma 374-0193,
            Japan (E-mail:fukumori@itakura.toyo.ac.jp, Tel:0276-82-9212,
            Fax:0276-82-9801)
COMMENT     On Jul 13, 2001 this sequence version replaced gi:1841358.
FEATURES             Location/Qualifiers
     source          1. .25404
                     /organism="Pseudomonas putida"
                     /plasmid="pTDN1"
                     /strain="UCC22"
                     /db_xref="taxon:303"
     gene            1. .1703
                     /gene="tnpA"
     CDS             <1. .1703
                     /gene="tnpA"
                     /codon_start=3
                     /transl_table=11
                     /protein_id="BAA12804.1"
                     /db_xref="GI:1841359"
                     /translation="ALYLKGSRKLPVEVVAPSLGMIWQVSISSPDRERAFQALEVATL
                     FALRRAVRNGSVWIEHSLSFRGRARLFFTDERWQAESKKHYARLSLPSKAATFLKPLL
                     ARVTAGVDAVAAAARSGVLRVDDELHLSPLPAEDEDPEVTKLRAALDHRIGEVQLPEV
                     ILAVDAQVRFSWIMLGREPRSTDELLMVYAGIMAHGTSLTAVECARMIPQLSATSIRQ
                     AMRWARDERRLSQACQAVLEFMQRHPIAATWGRSDLASSDMMTMETTKRVWQARLDPR
                     RNTPSIGIYSHVKDRWGIFHAQPFVLNERQAGVAIEGVIRQEKLETSQLAVDTHGYTD
                     FAMSHARLLGFDLCPRLKELKQRHLFVPRGTKVPAEIAAVCEANVDVALIEKHWDSLV
                     HLAASVMSGHASAVAALARFGSAAQGDPIYEAGVQLGRLLRTAFLADYFVKDAFRNEL
                     RRVLNRGEAVNALKRAIYTGRISPAQAKRVDEMQAVADALSLMANIVMAWNTSQMQAV
                     LDRWSNRRQVIPPELIGKIAPTRLESINLRGVFRFPVDRYADQILPSRPNASITGTNG
                     "
     gene            2802. .4280
                     /gene="tdnQ"
     CDS             2802. .4280
                     /gene="tdnQ"
                     /note="amino group transfer"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA12805.1"
                     /db_xref="GI:1841360"
                     /translation="MSGKFIEKHGIWSDTQKAAAADVLNKIEKAGLQMVRLSWPDQYG
                     LLRGKMLSVAALRSAFASGSEITMAPFFFDTASAIVFNPFSADGGLGSAELAGSPNVV
                     MVPDPTTFRILPWADRTGWMLADLYMTSGRPFALSPRAILKKALVEMQDLGYDYQAGL
                     EVEWYLTRIVDPCLEPETLGGPGTPAAPPKVMPVAKGYSYLLENHLDEVEPIMAEVRQ
                     HLLALGMPLRSIEDEWAPSQMETTFDVMPGLDVADTMVLFRNAVKQVCRRRGYLASFM
                     CKPAIQGFLASGWHLHQSLTARDSGANAFIPQPGEALSALGRSYVGGLLEHACAASSF
                     TTPTINGYRRRRPYSLAPDRVTWAKDNRAAMARVISAPGDPASRVENRIGEPAANPYL
                     YLASQVFSGIDGIRRQLDPGPLQETPYAGDVTILPHNLSEALEVLETSKFFREAFGEE
                     FIRYWMHLRRSEWKRFVDAEGQVDFSGDPVTNWEHREYFELF"
     gene            4399. .5160
                     /gene="tdnT"
     CDS             4399. .5160
                     /gene="tdnT"
                     /note="amino group transfer"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA12806.1"
                     /db_xref="GI:1841361"
                     /translation="MTAMKKYAVIWCSDASGDLELQEKMISAFGRENEEWEIIQPAEN
```

```
                     NFLERVFDYSGHVISGSPKSVIDDAQTPLVSNLLAFLRGAAQRGEGPVVGLCFGAQAI
                     AAALGGEVGRNPSGRFKLGVDRLEWSNEAQVLFGPQVGAGPTVLVQSHGECVTTLPPG
                     GVQLASSQTIPHEVFLVNGQFLGIQGHPEVDRQFLQQKFMAYHRALFDDDEWVGVQQE
                     SQQDLDPERVIALGRRLLDAGRLPATPQDISALPA"
gene                 5195. .6541
                     /gene="tdnA1"
CDS                  5195. .6541
                     /gene="tdnA1"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="large subunit of terminal dioxygenase"
                     /protein_id="BAA12807.1"
                     /db_xref="GI:1841362"
                     /translation="MTTTAKISVLPRTPAAQGDWDDLVQEDRVHRRLYTDEAIFSREM
                     NNIFAATWVYLAHESEIPEPNDFKQAWIGVREVIVARDEEGVIRVFSNRCSHRGASVC
                     REHRGNAAGFTCPYHGWRFDNRGQLFGIPGKNAYGPTFKSRDMHLARPAQVASYKGFV
                     FATLNPDAPSLAEHLGNAAPYLDAWIDHNGGPENLRLAGAQRFRLACNWKLVWDNAGD
                     GYHVPFSHQSLLVMTNERYGGGDMAYFADADRSKMTNSALDNGHTVIDQRPEMHGESA
                     WRLQRPQPGREPYEDHVASIYGDQAQQVLDTTVGAGMNLNIFPNLALIGNQVQVIQPL
                     AVASTGVHWYATQRKDADPEVNTMRLRTQEDFPVMGEMDDAANFEECQRGLCNSPEDE
                     WVDMSRHHESGKDVPVEDGIIRGPVTTDLHMRNYYAQWKRLMQAEPTLRMDKGRLA"
gene                 6538. .7182
                     /gene="tdnA2"
CDS                  6538. .7182
                     /gene="tdnA2"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="small subunit of terminal dioxygenase"
                     /protein_id="BAA12808.1"
                     /db_xref="GI:1841363"
                     /translation="MSTATQQPAAGRQGYRHQPPSLYVVASFYDWLVDVSADLARAVV
                     REPGTVEAARERDIVRMLTVEARLLDQGALTQDAYVQWLALFAEECAYWIPAVSPAPD
                     PRCSVTLEFHDRRRLLDRVTRLGTGLAFSQFPTSRTARQFSGLEVWASPGRSDEWRAR
                     YSFTLVESREGHGRVLAGWNGFVLRETEAGLSIVLKQVNLIDSDRPQGNNSFFL"
gene                 7194. .8207
                     /gene="tdnB"
CDS                  7194. .8207
                     /gene="tdnB"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="electron transfer protein"
                     /protein_id="BAA12809.1"
                     /db_xref="GI:1841364"
                     /translation="MAKRAQSLTITDVTAQGSDAILLSLRVDDEQQPKFTFQPGQYLT
                     LAVEVQGDEHWRCYSITSEPVTGQPISVLVRRVAGGRVSNWLCDNARPGLQLQVLPPA
                     GHFTLARPGQPLLLYAGGSGIAPVFALAREALARGAARVRLFYANRDRATAMLLAELQ
                     ALQDAAAGRLEIVHWYDAEQGLPTQAVLVAQAQGLDQADAYMCGPEPFMHAVGASLQV
                     AGFDAQRVHREDFGAAVEGASEEAAGDGPEALLTVLMKGQTHAVPVRAGELLLSAMLR
                     AGLPAPHACRVGECASCMCRLQAGEVQRLDSSVLDEDDVAAGWLLACRTRAASPALQV
                     RFS"
gene                 8255. .9142
                     /gene="tdnR"
CDS                  8255. .9142
                     /gene="tdnR"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="LysR-type regulator"
                     /protein_id="BAA12810.1"
                     /db_xref="GI:1841365"
                     /translation="MTRDGPTTPQWDLIRAFLALERHGSYEVAAELEGIDDSTLRRRI
                     RLLEQHFGRALFVRAEGGWRASADLNALISAAQRMEEAARSFCQDHHAGAGVIRISVM
                     DVFAQRFAPVFAALSEKYPKLVFTITTEAHFVNLEQDQVDIAVRLARPERNSNALRVR
                     KLGDVAVGAYASDAYLRRAAELAHENHQLLAMNLQFFHQDHHFIYASLDWSRFGLSGQ
                     VRFQSDSFAPLAQLCALGQGLALLPKFLAAEYPQLVAHPSNVFVDTQLWLVSRFDMHA
                     AWQRDLADMLQAEMARWPQ"
gene                 9154. .9513
                     /gene="tdnD1"
CDS                  9154. .9513
                     /gene="tdnD1"
                     /note="TdnD1"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="plant-type ferredoxin"
                     /protein_id="BAB62044.1"
                     /db_xref="GI:14715442"
                     /translation="MQATAVLGPSSVCVVQTGETYACSPGESLLRGMLRLGRKGIPVG
                     CVNGGCGVCKVRIVEGCVRALGPVSRAHVSTEEEAQGYTLACRVAPTQAVCLEVAARL
                     RKPFLGSAASVAQRPAN"
gene                 9550. .10494
                     /gene="tdnC"
CDS                  9550. .10494
                     /gene="tdnC"
                     /note="TdnC"
                     /codon_start=1
                     /transl_table=11
                     /product="catechol 2,3-dioxygenase"
                     /protein_id="BAB62045.1"
                     /db_xref="GI:14715443"
                     /translation="MGVLRIGHASLRVMDIAAAVKHYEEVLGLKTVMKDSAGNVYLKC
                     WDEWDKYSIILTPSDRAGLNHVAYKVEKDEDLEALQARIEAWGIKTTLLPEGTLPSTG
                     RMLQFNLPSGHEMRLYAMKEYVGTDVGTVNPDPWPDGVRGAGAHWLDHLLLMCEMNPE
                     AGINTVQDNTRFMKECLDFFLTEQVLVGPQGDMQAATWMARTSTPHDIAFVGGPVSGL
                     HHIAFFLDSWHDVLKSADVMAKKKVKIDVAPTRHGITRGETIYFFDPSGNRNETFAGL
                     GYLAQPDRPVTTWSEDKMGRGIFYHTGDLVPSFTDVYT".
gene                 complement(10485. .12083)
                     /gene="orf2"
CDS                  complement(10485. .12083)
                     /gene="orf2"
                     /note="ORF2
                     unknown product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB62046.1"
                     /db_xref="GI:14715444"
                     /translation="MDLRDLRHFLAVAEEGHIGRAAARLHLSQPPLTRHIQALEEKIG
                     VPLFVRTPKGMQLTDAGHTLLQETPNLLALAQQALERTRLAGQGLIGQLDVGLFGSGV
                     LDVIPRILARFHEERPEVKIVLHNLTKDAQLQALRERRISVGFNRLVPQEADISIETV
                     LHEPLVVALAASHPLAVRPSLRLSDLEGMPLILYPNVPLRGLAQEVTDAFRAEGIPPR
                     VEQEVQDVVTAVALVAGGFGCCITTRSATSLRLPGVTYRPLRSRHLSDIELSCLYRSD
                     VGAGARPAVFAGLPLAVAHQPRLPGAEGLAHPPLHRELIARAQAHFVLAVALETHQGR

  Query Match             72.8%;  Score 18.2;  DB 1;  Length 25404;
  Best Local Similarity   87.0%;  Pred. No. 6.2e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcg 23
           |||||| |||||||||| || |
Db    8837 GCAGGTTCATCGCCAGCAGTTGG 8815


RESULT  39
MTCY22D7/c
LOCUS       MTCY22D7    31859 bp    DNA             BCT       03-AUG-2001
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 133/162.
ACCESSION   Z83866 AL123456
VERSION     Z83866.1  GI:3261691
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis H37Rv.
  ORGANISM  Mycobacterium tuberculosis H37Rv
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 31859)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
```

```
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature. 393 (6685), 537-544 (1998)
  MEDLINE   98295987
REFERENCE   2  (bases 1 to 31859)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 27, 1998 this sequence version replaced gi:1781132.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES            Location/Qualifiers
     source         1. .31859
                    /organism="Mycobacterium tuberculosis H37Rv"
                    /strain="H37Rv"
                    /db_xref="taxon:83332"
     source         <1. .29946
                    /organism="Mycobacterium tuberculosis H37Rv"
                    /strain="H37Rv"
                    /db_xref="taxon:83332"
                    /clone="Y22D7"
     gene           complement(52. .504)
                    /gene="nrdI"
     CDS            complement(52. .504)
                    /gene="nrdI"
                    /note="Rv3052c, (MTCY22D7.30), len: 150. nrdI. Probable
                    nrdI protein, similar to many e.g. sp|Q47415|NRDI_ECOLI
                    NRDI PROTEIN (136 aa) fasta scores, opt: 574 z-score:
                    747.2 E(): 0; 62.2% identity in 135 aa overlap"
                    /codon_start=1
                    /transl_table=11
                    /product="nrdI"
                    /protein_id="CAB06244.1"
                    /db_xref="GI:1781162"
                    /db_xref="SWISS-PROT:P95107"
                    /translation="MDIAGRSLVYFSSVSENTHRFVQKLGIPATRIPLHGRIEVDEPY
                    VLILPTYGGGRANPGLDAGGYVPKQVIAFLNNDHNRAQLRGVIAAGNTNFGAEFCYAG
                    DVVSRKCSVPYLYRFELMGTEDDVAAVRTGLAEFWKEQTCHQPSLQSL"
     RBS            complement(518. .522)
                    /note="possible RBS, AAGGA, for Rv3052c"
     gene           complement(539. .778)
                    /gene="nrdH"
     CDS            complement(539. .778)
                    /gene="nrdH"
                    /note="Rv3053c, (MTCY22D7.29), len: 79. nrdH. Probable
                    NrdH protein, involved in electron transfer, shows strong
                    similarity to other NrdH glutaredoxin-like proteins. FASTA
                    best: sp|Q56108|NRDH_SALTY GLUTAREDOXIN-LIKE PROTEIN NRDH.
                    (81 aa) fatsa scores,   opt:  243 z-score: 358.1 E():
                    2e-12; 45.8% identity in 72 aa overlap"
                    /codon_start=1
                    /transl_table=11
                    /product="nrdH"
                    /protein_id="CAB06243.1"
                    /db_xref="GI:1781161"
                    /db_xref="SPTREMBL:P95106"
                    /translation="MTVTVYTKPACVQCSATSKALDKQGIAYQKVDISLDSEARDYVM
                    ALGYLQAPVVVAGNDHWSGFRPDRIKALAGAALTA"
     RBS            complement(789. .794)
                    /note="possible RBS, AAGGAG, for Rv3053c"
     misc_feature   complement(1004. .1017)
                    /note="conflict: this 14bp sequence is duplicated in BAC
                    and w.g.s. data"
     gene           complement(1255. .1809)
                    /gene="Rv3054c"
     CDS            complement(1255. .1809)
                    /gene="Rv3054c"
                    /note="Rv3054c, (MTCY22D7.28), len: 184. Function:
                    unknownbut shows some similarity to hypothetical bacterial
                    proteins and contains aminotransferases class-II
                    pyridoxal-phosphate attachment site (PS00599) near
                    C-terminus. FASTA best: YIEF_ECOLI P31465 hypothetical
                    20.4 kd protein in tnab- (188aa) opt:  324, E(): 1.3e-14,
                    (39.5% identity in 129 aa overlap), TBparse score is
                    0.873"
                    /codon_start=1
                    /transl_table=11
                    /product="hypothetical protein Rv3054c"
                    /protein_id="CAB06242.1"
                    /db_xref="GI:1781160"
                    /db_xref="SPTREMBL:P95105"
                    /translation="MSDTKSDIKILALVGSLRAASFNRQIAELAAKVAPDGVTVTMFE
                    GLGDLPFYNEDIDTATEVPAPVSALREAASDAHAALVVTPEYNGSIPAVIKNAIDWLS
                    RPFGDGALKDKPLAVIGGSMGRYGGVWAHDETRKSFSIAGTRVVDAIKLSVPFQTLGK
                    SVADDAGLAANVRDAVGNLAAEVG"
     misc_feature   complement(1312. .1341)
                    /gene="Rv3054c"
                    /note="PS00599 Aminotransferases class-II
                    pyridoxal-phosphate attachment site"
     RBS            complement(1818. .1822)
                    /note="possible RBS, AAGGA, for Rv3054c"
     gene           1901. .2515
                    /gene="Rv3055"
     CDS            1901. .2515
                    /gene="Rv3055"
                    /note="Rv3055, (MTCY22D7.26c), len: 204. Possible
                    regulatory protein, low similarity to several. FASTA best:
                    G987091 regulator of antibiotic transport complexes (204
                    aa)   90 opt:  150 z-score: 186.8 E(): 0.002, (28.9%
                    identity in 197aa overlap). Contains potential
                    helix-turn-helix motif fromaa 49to70(+3.29 SD)"
                    /codon_start=1
                    /transl_table=11
                    /product="hypothetical protein Rv3055"
                    /protein_id="CAB06241.1"
                    /db_xref="GI:1781158"
                    /db_xref="SPTREMBL:P95103"
                    /translation="MSGAERLGDLPVFARQEPVPERGDAARNRALLLEAARRLIARSG
                    ADAITMDDVAAAAGVGKGTLFRRFGSRAGLMMVLLDEDERASQQAFLFGPPPLGPDAP
                    PLDRLIAFGRERMRFVHAHHQLLSEANRDPQTRHSAALSVLRTHLRVLLASAPTTGDL
                    DAQTDALALLALLDVDYVEHQLNAGGHTLQTLGDAWESLARKLCGR"
     gene           2525. .3565
                    /gene="dinP"
     CDS            2525. .3565
                    /gene="dinP"
                    /note="Rv3056, (MTCY22D7.25c), len: 346. dinP, extensive
                    similarity to proteins induced by DNA damage such as dinP,
                    mucB, umuC. Also similar to MTCY48.28c. FASTA best:
                    YQJH_BACSU P54545 hypothetical 47.0 kd protein in glnq-
                    (414 aa) opt: 493, E(): 2e-26, (32.9% identity in 346 aa
                    overlap), TBparse score is 0.933"
                    /codon_start=1
                    /transl_table=11
                    /product="dinP"
                    /protein_id="CAB06240.1"
                    /db_xref="GI:1781157"
                    /db_xref="SPTREMBL:P95102"
```

```
                     /translation="MPTAAPRWILHVDLDQFLASVELLRHPELAGLPVIVGGNGDPTE
                     PRKVVTCASYEARAYGVRAGMPLRTAARRCPEATFLPSNPAAYNAASEEVVALLRDLG
                     YPVEVWGWDEAYLAVAPGTPDDPIEVAEEIRKVILSQTGLSCSIGISDNKQRAKIATG
                     LAKPAGIYQLTDANWMAIMGDRTVEALWGVGPKTTKRLAKLGINTVYQLAHTDSGLLM
                     STFGPRTALWLLLAKGGGDTEVSAQAWVPRSRSHAVTFPRDLTCRSEMESAVTELAQR
                     TLNEVVASSRTVTRVAVTVRTATFYTRTKIRKLQAPSTDPDVITAAARHVLDLFELDR
                     PVRLLGVRLELA"
     gene            complement(3619. .4482)
                     /gene="Rv3057c"
     CDS             complement(3619. .4482)
                     /gene="Rv3057c"
                     /note="Rv3057c, (MTCY22D7.24), len: 287. Function:
                     unknownbut similar to various oxidoreductases and enzymes
                     involvedin polyketide synthesis. Contains short-chain
                     alcohol dehydrogenase family signature (PS00061) and is
                     related to otherM. tuberculosis proteins e.g. MTCY190.25c,
                     MTCY39.16c. FASTA best: Y00P_MYCTU Q10402 putative
                     oxidoreductase cy190.25c (650 aa) opt: 462, E(): 3.9e-22;
                     (34.5% identity in 258aaoverlap). TBparse score is 0.925"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv3057c"
                     /protein_id="CAB06265.1"
                     /db_xref="GI:1781156"
                     /db_xref="SPTREMBL:P95101"
                     /translation="MLQRGAGQYFAGKRCFVTGAASGIGRATALRLAAQGAELYLTDR
                     DRDGLAQTVCDARALGAQVPEHRVLDVSDYQDVAAFAADIHARHPSMDVVLNIAGVSA
                     WGTVDQLTHDQWSRMVAINLMGPIHVIETLVPPMVAAGRGGHLVNVSSAAGLVGLPWH
                     AAYSASKYGLRGLSEVLRFDLARHGIGVSVVVPGAVKTPLVNTVEIAGVDRDDPRVNR
                     WVERFSGHAVTPEKAADKILAGVTRNRYLVYTSADIRALYAFKRYAWWPYTLVMRRVN
                     VFFTRALRPGP"
     misc_feature    complement(3949. .4035)
                     /gene="Rv3057c"
                     /note="PS00061 Short-chain alcohol dehydrogenase family
                     signature"

  Query Match             72.8%;  Score 18.2;  DB 1;  Length 31859;
  Best Local Similarity   87.0%;  Pred. No. 6.2e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcga 24
           |||||||| |||||||| | |||
Db   19871 CAGGTGCACCGCCAGCAATCCGA 19849


RESULT  40
AC014955/c
LOCUS       AC014955   103920 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC014955
VERSION     AC014955.1  GI:6436380
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 103920)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212515 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .103920
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    29949 a  21507 c  21794 g  30670 t
ORIGIN

  Query Match             72.8%;  Score 18.2;  DB 2;  Length 103920;
  Best Local Similarity   87.0%;  Pred. No. 6e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 aggtgcatcgccagcatttcgat 25
           ||| || |||||||||||| ||||
Db   66862 AGGGGCTTCGCCAGCATTCCGAT 66840


RESULT  41
AC091510/c
LOCUS       AC091510   104000 bp    DNA             HTG       25-JUL-2001
DEFINITION  Leishmania major chromosome Unknown clone P1023 strain Friedlin,
            *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.
ACCESSION   AC091510
VERSION     AC091510.2  GI:15011696
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Leishmania major.
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 104000)
  AUTHORS   Myler,P.J., Sisk,E., Cawthra,J., Nelson,S., Vogt,C., Robertson,L.,
            McDonagh,P., Ivens,A., Mack,J., Munden,H., Stuart,K., Marty,A. and
            Seyler,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2001) Seattle Biomedical Research Institution, 4
            Nickerson Street, Seattle, WA 98109-1651, USA
COMMENT     On Jul 25, 2001 this sequence version replaced gi:13811912.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    19381: contig of 19381 bp in length
            *    19382    19525: gap of unknown length
            *    19526   104000: contig of 84475 bp in length.
            This PAC clone appears to not contain Leishmania DNA and may
            represent contaminating bacterial genomic DNA.
FEATURES             Location/Qualifiers
     source          1. .104000
                     /organism="Leishmania major"
                     /strain="Friedlin"
                     /db_xref="taxon:5664"
                     /chromosome="Unknown"
                     /clone="P1023"
BASE COUNT    16290 a  35980 c  35169 g  16417 t    144 others
ORIGIN

  Query Match             72.8%;  Score 18.2;  DB 2;  Length 104000;
  Best Local Similarity   87.0%;  Pred. No. 6e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcg 23
           |||||||| ||||||||| |||
Db   77937 GCAGGTGCCGCGCCAGCATGTCG 77915


RESULT  42
AE003544/c
LOCUS       AE003544   281869 bp    DNA             INV       06-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386050 section 24
            of 54, complete sequence.
ACCESSION   AE003544 AE002602
```

```
VERSION     AE003544.2  GI:10727982
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 281869)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 281869)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7294680.
FEATURES             Location/Qualifiers
     source          1. .281869
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="3L"
     mRNA            complement(5394. .>6614)
                     /gene="CG12289"
                     /product="CT19318"
                     /db_xref="FLYBASE:FBan0012289"
                     /db_xref="FLYBASE:FBgn0036160"
     gene            complement(<5394. .>6614)
                     /gene="CG12289"
                     /map="68C3-68C3"
                     /db_xref="FLYBASE:FBan0012289"
                     /db_xref="FLYBASE:FBgn0036160"
     CDS             complement(5394. .6503)
                     /gene="CG12289"
                     /note="CG12289 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0012289"
                     /db_xref="FLYBASE:FBgn0036160"
                     /protein_id="AAF50065.1"
                     /db_xref="GI:7294728"
                     /translation="MKAVLCVGCTEVDYVTTMDKVDFQCDRCSTQNGMLQRSGRSSNV
                     STALRLLGANVELFGVLTRNPTYRMILDDLEQRGIDISHCTFTDASPPFSTIVQDRWT
                     RRCTVVYCDKSFPYVTAADFRKLDLDQYGWVNFEARSPRETCDMIRLILDHNNGRDER
                     DRIVVSLQIFNSFEQVVHLIAMCDYVFVTKYIAESRGWKSPLEACEGVNALLRMPRNI
                     RIRRPCIIVPWSSLGAGCMTTEGEYFEVPAHKPKKIIDRVGDSDCFTAGFIYAAFVRQ
                     RRLAEAVDFANAVASYKLGYVGFDCLGHLPIDAVKLPVRMRCGSEVSSDDEDNGDQLN
                     SCKKHLLRPIEFLPEKSSIAPTEKEVPIETSQLSS"
     mRNA            join(<6798. .7180,7307. .7723,7807. .>7834)
                     /gene="CG7551"
                     /product="CT23115"
                     /db_xref="FLYBASE:FBan0007551"
                     /db_xref="FLYBASE:FBgn0036161"
     gene            <6798. .>7834
                     /gene="CG7551"
                     /map="68C3-68C3"
                     /db_xref="FLYBASE:FBan0007551"
                     /db_xref="FLYBASE:FBgn0036161"
     CDS             join(6798. .7180,7307. .7723,7807. .7834)
                     /gene="CG7551"
                     /note="CG7551 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0007551"
                     /db_xref="FLYBASE:FBgn0036161"
                     /protein_id="AAF50064.1"
                     /db_xref="GI:7294727"
                     /translation="MTQYGHKLSWAVGKKTVLCIGTTTIDFVSTIKRFPEENSHQKVI
                     GGYWIRGGKASNNCTVLANLGVKVEFLGMLSSWSVFQVLVDDLKSRGIIIDNCPTCDQ
                     GAPFSSVILTKATKTRNIVYCNNNFPKEKIVLSMEFDNNLDDMWPLMDYCDYAVFSKK
                     LAQPNGWVSLEDACMQLDERLRMRWGLNLKRPYVVVLWGDQGAGILDLNGNFTHVKPH
                     KPKRIVDALGAGDTFVGAFIYALYIRERSVSVAVDFGNRMASYKCTKNGKQIKRMRL"
     mRNA            complement(join(<12006. .12265,12321. .13329,13402. .>13545))
                     /gene="CG6140"
                     /product="CT19294"
                     /db_xref="FLYBASE:FBan0006140"
                     /db_xref="FLYBASE:FBgn0036162"
                     /evidence=not_experimental
     gene            complement(<12006. .>13545)
                     /gene="CG6140"
                     /map="68C3-68C3"
                     /db_xref="FLYBASE:FBan0006140"
                     /db_xref="FLYBASE:FBgn0036162"
                     /evidence=not_experimental
     CDS             complement(join(12006. .12265,12321. .13329,13402. .13545))
                     /gene="CG6140"
                     /note="CG6140 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0006140"
                     /db_xref="FLYBASE:FBgn0036162"
                     /evidence=not_experimental
                     /protein_id="AAF50063.1"
                     /db_xref="GI:7294726"
                     /translation="MSSTKKVETRQESDTLGPMEVPMDRYYGAQTMRCLINFRIGGEE
                     ERMPRQIIQAMGILKKAAAETNQEFGLDPKLSTAISNAADDVISGKLYDEGHFPLPIW
                     QTGSGTQSNMNSNEVIGNRAIELLGGRIGTKDPVHPNDHVNKSQSSNDTFPSAIHIAV
                     ATALTKDLRPAVTALRDSLQAKSNEWKDIIKIGRTHTQDAVPLTLGQEFSGYAQQLTN
                     GLQRIDAVLPRVYQLALGGTAVGTGLNTRRGFAEKCVKRIAQLSGLPFVVAPNFFEAL
                     ACRDAMVEVHGALNVLAVSLMKVTNDIRFLGSGPRCGLGELFLPENEPGSSIMPGKVN
                     PTQCEAMTMICAQVMGNHVAVSVGGANGHFELNVFKPLIASNVLRSIKLLADGCISFN
                     CNCVKGIKPNKEKLAKIVNESLMLVTALNPHIGYDKSAQIAKAAHKNGTTLKVEALNA
                     GISEKDFNEWVRPEKMLGPS"
     mRNA            join(<21646. .22118,23232. .>23712)
                     /gene="CG7547"
                     /product="CT23105"
                     /db_xref="FLYBASE:FBan0007547"
                     /db_xref="FLYBASE:FBgn0036163"
     gene            <21646. .>23712
                     /gene="CG7547"
                     /map="68C3-68C4"
```

```
                     /db_xref="FLYBASE:FBan0007547"
                     /db_xref="FLYBASE:FBgn0036163"
     CDS             join(21646. .22118,23232. .23712)
                     /gene="CG7547"
                     /note="CG7547 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0007547"
                     /db_xref="FLYBASE:FBgn0036163"
                     /protein_id="AAF50062.1"
                     /db_xref="GI:7294725"
                     /translation="MATDAADKDKPAYSFFFGSMGAASAIIFSALGAAYGTAKSGTGI
                     AAMAVMRPELIMKSIIPVVMAGIIAIYGLVVSVLIAGSLSDSYTIRKGYIHLAAGLSV
                     GFAGLAAGFAIGIVGDAGVRGTAQQPRLFVGMILILIFAEVLGLYGLIVAIYLYTNKM
                     VTAALNEEPSYAFFLGCTGAAVAIIFTTLGASYGTAVSGVGIAKMAVNRPDMIMKAII
                     PVVMAGIIAIYGLVVSVLIAGSIGDDYTMEDSYVHLGAGLSVGLPGLTAGVAIGIAGD
                     AGVRGTAEQPRLFVGMVLILIFAEVLALYGLIVAIYLYTKL"
     mRNA            join(<26812. .26991,28979. .29261,30903. .31335,32541. .32677,
                     33043. .33273,33411. .33672,35647. .35761,35930. .>36106)
                     /gene="CG7543"
                     /product="CT23091"
                     /db_xref="FLYBASE:FBan0007543"
                     /db_xref="FLYBASE:FBgn0036164"
     gene            <26812. .>36106
                     /gene="CG7543"
                     /map="68C3-68C4"
                     /db_xref="FLYBASE:FBan0007543"
                     /db_xref="FLYBASE:FBgn0036164"
     CDS             join(26812. .26991,28979. .29261,30903. .31335,32541. .32677,
                     33043. .33273,33411. .33672,35647. .35761,35930. .36106)
                     /gene="CG7543"
                     /note="CG7543 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0007543"
                     /db_xref="FLYBASE:FBgn0036164"
                     /protein_id="AAF50061.1"
                     /db_xref="GI:7294724"
                     /translation="MTPLDFYREQHLHTKNHVEPNLQSVKTVAASQSRTIRSKSNSFA
                     NRTEIMRRQERCGGPHQTVDPTQANRSQETQPNPVQLSSSSIWEPKHRKSFLELIEFK
                     RMKALNRFVFEGRRIDAELRPFSERKVVGVGKVKVRDTTLPLVGEGYSGSPITIAQEL
                     PLPQAAKQSNINRLWVVTGGIRGFAGFEGFVGLVLAEIAGFWVATGPLMCACRRHGKV
                     LEELHTMHALFMLTSEFALQSKKCCKDTCAMPGKSSWPPNGKQSRRMWSATKTRAGSV
                     NEKAGLRSTQRSRPKRKSAKYSRAARVLGGCRQRATWDGIGMGQDEMVRVANPNANAD
                     GKGAGAAVDAALGEQTKKPLSGSFSFLLPGDVSKNLKDCRPCCCSYFIGKTCNALEEP
                     APQFPPAPQRLTKWSIRRRRWRCQWPWLDERIASLSANGSRALVEAFSKQFLRLPFRI
                     TFVLAKCICVSANVNGFSSRLATCGKVSNHEKTSQLNGYAMWVLAAFMVEKLVLLAIL
                     GTADHRLQKSAKSLKREFTAKEKENEKLRRRQRSRRGKLLFPSSLARTNTETETETAE
                     RALSENFVVRFTRTSEKNRSRNAAAKHSTSAHKGQNEKN"
     mRNA            join(<36773. .36839,42328. .42970,43084. .43194,43256. .43751)
                     /gene="charybde"
                     /product="CT23077"

  Query Match             72.8%;  Score 18.2;  DB 3;  Length 281869;
  Best Local Similarity   87.0%;  Pred. No. 5.9e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 aggtgcatcgccagcatttcgat 25
           ||| || |||||||||||| ||||
Db 232720 AGGGGCTTCGCCAGCATTCCGAT 232698


RESULT  43
MAV250020
LOCUS       MAV250020    2923 bp    DNA             BCT       26-SEP-2000
DEFINITION  Mycobacterium avium subsp. paratuberculosis insertion sequence
            IS900, Locus 7.
ACCESSION   AJ250020
VERSION     AJ250020.1  GI:8919146
KEYWORDS    2599 gene; 2600 gene; insertion element; insertion sequence IS900;
            p43 gene; transposase.
SOURCE      Mycobacterium avium subsp. paratuberculosis.
  ORGANISM  Mycobacterium avium subsp. paratuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium avium complex (MAC).
REFERENCE   1  (bases 1 to 2923)
  AUTHORS   Bull,T.J., Hermon-Taylor,J., Pavlik,I., El-Zaatari,F. and Tizard,M.
  TITLE     Characterization of IS900 loci in mycobacterium avium subsp.
            paratuberculosis and development of multiplex PCR typing
  JOURNAL   Microbiology 146 (Pt 9), 2185-2197 (2000)
  MEDLINE   20431891
REFERENCE   2  (bases 1 to 2923)
  AUTHORS   Bull,T.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-1999) Bull T.J., Surgery, St.George's Hospital
            Medical School, London, SW17 ORE, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2923
                     /organism="Mycobacterium avium subsp. paratuberculosis"
                     /sub_species="paratuberculosis"
                     /db_xref="taxon:1770"
     gene            complement(11. .466)
                     /gene="2600"
     CDS             complement(11. .466)
                     /gene="2600"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB96064.1"
                     /db_xref="GI:8919147"
                     /translation="MYLAVEIGTVDLNPVLKGAVATILYFAVGMAVLLVGFYAVDLLT
                     PGKLRQLVFIDRRPNAVVVAGAMYIALTVVIITAIANSYSQLGQGLVGVAVYGLMGVI
                     LLGVALLAMHLLIPGSFHEHVEEPQLHPGSFAVALILLAVGGVTAAAVS"
     source          473. .1922
                     /organism="Mycobacterium avium subsp. paratuberculosis"
                     /insertion_seq="IS900"
                     /sub_species="paratuberculosis"
                     /db_xref="taxon:1770"
     gene            686. .1906
                     /gene="p43"
     CDS             686. .1906
                     /gene="p43"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="CAB96065.1"
                     /db_xref="GI:8919148"
                     /translation="MTVTEVVVAQPVWAGVDAGKADHYCMVINDDAQRLLSQRVANDE
                     AALLELIAAVTTLADGGEVTWAIDLNAGGAALLIALLIAAGQRLLYIPGRTVHHAAGS
                     YRGEGKTDAKDAAIIADQARMRHDLQPLRAGDDIAVELRILTSRRSDLVADRTRAIEP
                     NARPAAGILSALERAFDYNKSRAALILLTGYQTPDALRSAGGARVAAFLRKRKARNAD
                     TVAATALQAANAQHSIVPGQQLAATVVARLAKEVMALDTEIGDTDAMIEERFRRHRHA
                     EIILSMPGFGVILGAEFLAATGGDMAAFASADRLAGVAGLAPVPRDSGRISGNLKRPR
                     RYDRRLLRACYLSALVSIRTDPSSRTYYDRKRTEGKRHTQAVLALARRRLNVLWAMLR
                     DHAVYHPATTTAAA"
     gene            complement(1963. .2394)
                     /gene="2599"
     CDS             complement(1963. .>2394)
                     /gene="2599"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB96066.1"
                     /db_xref="GI:8919149"
                     /translation="VSRNRLFVIAGALAVAAVVCLVTGIILLQKNIASYVAGHYHEYA
                     RDVNGTRYQCTGSSEQVADTLSEYAEPAARADDNKTEYLRYSNDIVIVGPDGNYPCSI
                     RVEPLSAGYSHRAFVFLGPGFFPGSPSGGAGGTPGGPGGTK"
BASE COUNT      473 a   1005 c   1000 g    445 t
ORIGIN


  Query Match             72.0%;  Score 18;  DB 1;  Length 2923;
  Best Local Similarity   100.0%;  Pred. No. 8.1e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 gcaggtgcatcgccagca 18
           ||||||||||||||||||
Db     127 GCAGGTGCATCGCCAGCA 144


RESULT  44
G01446
LOCUS       G01446        479 bp    DNA             STS       19-AUG-1999
DEFINITION  Dm0466 Drosophila P1 library Drosophila melanogaster STS genomic
            clone DS07967 SP6, sequence tagged site.
ACCESSION   G01446
VERSION     G01446.1  GI:684849
KEYWORDS    STS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   Rubin,G.
  TITLE     Berkeley Drosophila Genome Project
  JOURNAL   Unpublished (1994)
COMMENT
            Contact:
            Berkeley Drosophila Genome Project
            Primer A: CTCTCACTGGGGACTTGCTA
            Primer B: CACAAAGGGCATATCCGA
            STS size: 192
            PCR Profile:
                 Annealing:  58 degrees C      PCR Cycles: 30
            Protocol:
                 Template:    P1 Library Pools
                 Primer:      1 uM each
                 dNTPs:       250 uM each
                 Taq Poly:    0.05 units/ul
                 Total Vol:   15 ul

            Buffer:
                 MgCl2:       1.5mM
                 KCl:        50  mM
                 Tris-HCl:   50  mM
                 pH:          8.3
                 Gelatin:      .001 %

            The P1 library has been distributed to 16 regional sites.  A list
            of these sites is available from FlyBase, via anonymous ftp to
            ftp.bio.indiana.edu in the file
            flybase/allied-data/genome-projects/lbl/LBL.doc.
FEATURES             Location/Qualifiers
     source          1. .479
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="DS07967"
                     /clone_lib="Drosophila P1 library"
                     /note="Vector: PAd10SacII; The P1 library was made by D.
                     Smoller in D. Hartl's lab (see Smoller et al., Chromosoma
                     100: 487).  Chromosomal position was mapped by polytene
                     chromosome in situ hybridization in the Hartl lab (see
                     Hartl et al., PNAS 91: 6824).  STS's were generated by
                     sequencing the ends of the Drosophila insert in these P1
                     clones, by the W. Kimmerly, C. Martin, and M. Palazzolo
                     lab at LBL."
     STS             <1. .250
     primer_bind     complement(233. .250)
BASE COUNT      113 a    120 c     97 g    145 t      4 others
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 11;  Length 479;
  Best Local Similarity   90.5%;  Pred. No. 1e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 aggtgcatcgccagcatttcg 23
```

```
           | | ||||||||||||||||||
Db     392 ATGCGCATCGCCAGCATTTCG 412


RESULT  45
MSU31280
LOCUS       MSU31280     3748 bp    DNA             BCT       21-NOV-1996
DEFINITION  Methylobacillus sp. beta-lactamase (bla) and biotin synthase (bioB)
            genes, complete cds.
ACCESSION   U31280
VERSION     U31280.1  GI:1680651
KEYWORDS    .
SOURCE      Methylobacillus sp. KT1.
  ORGANISM  Methylobacillus sp. KT1
            Bacteria; Proteobacteria; beta subdivision; Methylophilus group;
            Methylobacillus.
REFERENCE   1  (bases 1 to 3748)
  AUTHORS   Serebriiskii,I.G., Vassin,V.M. and Tsygankov,Y.D.
  TITLE     Two new members of the bio B superfamily: cloning, sequencing and
            expression of bio B genes of Methylobacillus flagellatum and
            Corynebacterium glutamicum
  JOURNAL   Gene 175 (1-2), 15-22 (1996)
  MEDLINE   97074643
REFERENCE   2  (bases 1 to 3748)
  AUTHORS   Serebrijski,I., Vassin,V. and Tsygankov,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-1995) Ilya Serebrijski, Institute of Genetics and
            Selection of Industrial Microorganisms, 1st Dorozhny Proezd,
            Moscow, 113545, Russia
FEATURES             Location/Qualifiers
     source          1. .3748
                     /organism="Methylobacillus sp. KT1"
                     /isolate="KT1"
                     /db_xref="taxon:79675"
     gene            201. .1061
                     /gene="bla"
     CDS             201. .1061
                     /gene="bla"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-lactamase"
                     /protein_id="AAC44581.1"
                     /db_xref="GI:1680652"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
     -35_signal      2141. .2153
                     /gene="bioB"
     gene            2141. .3243
                     /gene="bioB"
     -10_signal      2177. .2183
                     /gene="bioB"
     RBS             2207. .2211
                     /gene="bioB"
     CDS             2218. .3243
                     /gene="bioB"
                     /note="Method: conceptual translation supplied by author"
                     /codon_start=1
                     /transl_table=11
                     /function="sulfur inserting enzyme; last step of biotin
                     biosynthesis"
                     /product="biotin synthetase"
                     /protein_id="AAC44582.1"
                     /db_xref="GI:1680653"
                     /translation="MTMITNSHRMRDSVIDIRGLDRVQRARVQAKEEAPKRWSVDDIV
                     ALFELPFSDLMHRAQSVHRENFDPNGVQVSTLLSIKTGGCSEDCGYCPQAARYHTDVE
                     KQDLHAAGRGTRSRVLPRRMRQPFCMAAWQPKQRDLEPVLAMIREVKAMGLETCATLG
                     MLKDGQAEQLKEAGLDYYNHNLDTAPEYYGEVITTRTYQDRLDTLDRVREQDINVCCG
                     GIIGMGESRVQRAGLLAQLANMERPPESVPDQLLTQVEGTPMYGMDELDPFEFVRTIA
                     AARITMPQSFVRLSVGRQSMHEGIQALCFLAGANSIFYGEKLLTTGNPEAEADRKLFD
```

KLGIHPL"
BASE COUNT 895 a 980 c 988 g 885 t
ORIGIN

Query Match 71.2%; Score 17.8; DB 1; Length 3748;
Best Local Similarity 90.5%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 gtgcatcgccagcatttcgat 25
||||||||| || ||||||||
Db 2398 GTGCATCGCGAGAATTTCGAT 2418

Search completed: December 26, 2001, 10:57:33
Job time: 2015 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:18:36 ; Search time 366.42 Seconds
(without alignments)
58.493 Million cell updates/sec

Title: US-09-396-196F-6
Perfect score: 25
Sequence: 1 gcaggtgcatcgccagcatttcgat 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	839	22	AAI62941	Human genomic DNA
	2	25	100.0	1041	20	AAX01303	E. coli biotin syn
	3	25	100.0	1084	10	AAN91329	E.coli Bio B gene.
	4	25	100.0	1121	7	AAN60496	Sequence encoding
	5	25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
	6	17.6	70.4	3084	18	AAT68959	Candida albicans c
	7	17.6	70.4	3617	17	AAT37079	AR301 fragment enc
c	8	17.4	69.6	4674	22	AAH18667	Human cDNA sequenc
c	9	17.2	68.8	3186	19	AAV39842	Pseudomonas fluore
c	10	17.2	68.8	3186	19	AAV31654	Nucleotide sequenc
c	11	17.2	68.8	3186	20	AAX99373	P. fluorescens lem
	12	17	68.0	669	21	AAZ53592	Neisseria gonorrhe
	13	17	68.0	759	21	AAZ53594	Neisseria meningit
	14	17	68.0	6386	20	AAX99364	Maize shrunken-1 (
c	15	17	68.0	23532	21	AAA81455	N. meningitidis pa
c	16	17	68.0	109519	22	AAS08693	Micromonospora DNA
c	17	17	68.0	349980	21	AAF21610	Neisseria meningit
c	18	17	68.0	1437668	21	AAA81490	N. meningitidis B
c	19	16.8	67.2	643	20	AAV69543	P. neopolydactyla
	20	16.6	66.4	65	18	AAT64649	G-CSF receptor ago
	21	16.6	66.4	522	18	AAT64616	G-CSF receptor ago
	22	16.6	66.4	531	18	AAT64612	G-CSF receptor ago
c	23	16.6	66.4	584	22	AAI42059	Probe #10745 used
	24	16.6	66.4	1066	22	AAS22818	Human cDNA encodin
	25	16.6	66.4	1255	22	AAS22582	Human cDNA encodin
	26	16.6	66.4	2074	20	AAX04354	Human secreted pro
c	27	16.6	66.4	2568	22	AAH64980	C glutamicum codin
c	28	16.6	66.4	4372	18	AAT91360	Orf virus genomic
c	29	16.6	66.4	10391	20	AAX13070	Enterococcus faeca
c	30	16.6	66.4	29879	14	AAQ46806	eryA region of S.
c	31	16.6	66.4	349980	22	AAH64966	C glutamicum codin
c	32	16.4	65.6	2048	22	AAF26295	Pseudomonas sp Typ
	33	16.2	64.8	645	21	AAZ54339	Neisseria gonorrhe
	34	16.2	64.8	645	21	AAZ54340	Neisseria meningit
	35	16.2	64.8	645	21	AAZ54341	Neisseria meningit
c	36	16.2	64.8	900	21	AAZ53354	Neisseria meningit
c	37	16.2	64.8	900	21	AAZ53355	Neisseria meningit
c	38	16.2	64.8	912	21	AAZ53353	Neisseria gonorrhe
c	39	16.2	64.8	1896	20	AAX77011	Partial Sel-1L cod
c	40	16.2	64.8	2109	20	AAX77014	Human Sel-1L codin
c	41	16.2	64.8	2144	22	AAF93859	Human cDNA encodin
c	42	16.2	64.8	2382	20	AAX76577	Human pancreas-spe
c	43	16.2	64.8	3666	20	AAX77013	Mouse Sel-1L splic
c	44	16.2	64.8	3667	20	AAX77012	Full length Mouse
c	45	16.2	64.8	5662	22	AAI61089	Human polynucleoti
c	46	16.2	64.8	5773	21	AAC76014	Human ORFX ORF1569
c	47	16.2	64.8	5876	22	AAI59303	Human polynucleoti
c	48	16.2	64.8	7885	20	AAX76578	Human pancreas-spe
c	49	16.2	64.8	8028	21	AAC75460	Human ORFX ORF1015
	50	16.2	64.8	9880	21	AAZ46651	Nucleotide sequenc
	51	16.2	64.8	47475	21	AAA81465	N. meningitidis pa
c	52	16.2	64.8	66986	22	AAF28542	Genomic fragment #
c	53	16.2	64.8	69936	21	AAA81479	N. meningitidis pa
	54	16.2	64.8	349980	21	AAF21607	Neisseria meningit
	55	16.2	64.8	349980	21	AAF21612	Neisseria meningit
	56	16.2	64.8	837096	21	AAA81489	N. meningitidis pa
	57	16.2	64.8	1437668	21	AAA81490	N. meningitidis B
c	58	16	64.0	199	22	AAI27192	Probe #17125 for g
c	59	16	64.0	199	22	AAI56045	Probe #24731 used
c	60	16	64.0	390	21	AAC74718	Human ORFX ORF273
c	61	16	64.0	396	21	AAH30338	Human colon cancer
	62	16	64.0	400	20	AAV90273	EST clone D1448.
	63	16	64.0	416	22	AAF93390	cDNA encoding SRT
	64	16	64.0	506	21	AAA42475	Human secreted exp
c	65	16	64.0	512	22	AAI27113	Probe #17046 for g
c	66	16	64.0	512	22	AAI55955	Probe #24641 used
c	67	16	64.0	531	21	AAC47430	Arabidopsis thalia
c	68	16	64.0	539	21	AAC36769	Arabidopsis thalia
c	69	16	64.0	554	22	AAI17977	Probe #7910 for ge
c	70	16	64.0	554	22	AAI42966	Probe #11652 used
c	71	16	64.0	575	22	AAI17923	Probe #7856 for ge
c	72	16	64.0	575	22	AAI42902	Probe #11588 used
	73	16	64.0	583	21	AAF09374	Fusarium venenatum
	74	16	64.0	583	22	AAH69466	Human cervical can
	75	16	64.0	664	20	AAZ40458	Human gonadotropin
c	76	16	64.0	732	22	AAH53836	S. epidermidis ope
	77	16	64.0	738	20	AAX39714	Gastric cancer ass
c	78	16	64.0	750	20	AAX32469	P. fluorescens mal
	79	16	64.0	764	20	AAZ40464	Human gonadotropin
c	80	16	64.0	784	21	AAC48286	Arabidopsis thalia
c	81	16	64.0	786	20	AAX39713	Gastric cancer ass
c	82	16	64.0	786	21	AAC37250	Arabidopsis thalia
	83	16	64.0	818	20	AAX39715	Gastric cancer ass
c	84	16	64.0	927	20	AAX39716	Gastric cancer ass

c	85	16	64.0	1020	21	AAA08593	Human cytoskeleton
c	86	16	64.0	1068	22	AAH53889	S. epidermidis ope
c	87	16	64.0	1135	21	AAC36656	Arabidopsis thalia
c	88	16	64.0	1341	22	AAF61024	P. putida KT2440-a
c	89	16	64.0	1423	21	AAC76080	Human ORFX ORF1635
c	90	16	64.0	1425	22	AAI60825	Human polynucleoti
	91	16	64.0	1479	22	AAH81414	Escherichia coli p
c	92	16	64.0	1491	22	AAH66345	C glutamicum codin
	93	16	64.0	1498	21	AAC51224	Arabidopsis thalia
	94	16	64.0	1500	21	AAC36258	Arabidopsis thalia
c	95	16	64.0	1809	22	AAF94396	Haemophilus influe
c	96	16	64.0	1812	20	AAX18925	Phaseolus lunatus
c	97	16	64.0	1997	22	AAI59039	Human polynucleoti
c	98	16	64.0	2263	19	AAV09991	DNA encoding a Hae
c	99	16	64.0	2534	12	AAQ10167	Cephalosporin C am
	100	16	64.0	2693	19	AAV71119	Exemplery caffeic

ALIGNMENTS

```
RESULT   1
AAI62941
ID   AAI62941 standard; DNA; 839 BP.
XX
AC   AAI62941;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human genomic DNA SEQ ID NO 269.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant;  immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200155449-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01346.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUL-2000; 2000US-0216880.
PR   14-JUL-2000; 2000US-0218290.
PR   14-AUG-2000; 2000US-0225447.
PR   01-SEP-2000; 2000US-0229343.
PR   06-SEP-2000; 2000US-0230437.
PR   08-SEP-2000; 2000US-0231243.
PR   25-SEP-2000; 2000US-0234997.
PR   29-SEP-2000; 2000US-0236367.
PR   13-OCT-2000; 2000US-0239937.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246528.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249265.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-476225/51.
XX
PT   Novel plasma membrane associated proteins useful for diagnosing,
PT   treating, preventing and/or prognosing disorders related to the
PT   proteins, including cancer, immune response and neuronal disorders  -
XX
PS   Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (AAI62752-AAI62961) and proteins
CC   (AAM42347-AAM42415) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;


  Query Match             100.0%;  Score 25;  DB 22;  Length 839;
  Best Local Similarity   100.0%;  Pred. No. 0.017;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||||||||||||||||||||
Db     116 gcaggtgcatcgccagcatttcgat 140


RESULT   2
AAX01303
ID   AAX01303 standard; DNA; 1041 BP.
XX
AC   AAX01303;
XX
DT   12-APR-1999  (first entry)
XX
DE   E. coli biotin synthetase (BioB) coding sequence.
XX
KW   DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW   biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS   Escherichia coli.
XX
PN   US5869719-A.
XX
PD   09-FEB-1999.
XX
PF   30-APR-1997;   97US-0846338.
XX
PR   30-APR-1997;   97US-0846338.
PR   08-MAR-1995;   95US-0401068.
XX
```

```
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Patton DA;
XX
DR   WPI; 1999-152902/13.
DR   P-PSDB; AAW73906.
XX
PT   Transgenic plants with high biotin levels - transformed with DNA
PT   encoding di:amino-pelargonic acid amino-transferase or biotin
PT   synthase
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.018;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcgat 25
          |||||||||||||||||||||||||
Db     84 gcaggtgcatcgccagcatttcgat 108


RESULT   3
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA;
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 0.018;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcgat 25
          |||||||||||||||||||||||||
Db    107 gcaggtgcatcgccagcatttcgat 131


RESULT   4
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzyme -
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 0.018;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcgat 25
          |||||||||||||||||||||||||
Db    125 gcaggtgcatcgccagcatttcgat 149


RESULT   5
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
```

```
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry)
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag= a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag= b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag= c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag= d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag= e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag= f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag= g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag= h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag= i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag= j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag= k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag= l
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag= m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag= n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                   7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag= o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag= p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag= q
FT                   /standard_name= "rho-independent transcriptional
FT                                   terminator"
FT   stem_loop       5583..5605
FT                   /*tag= r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;   93WO-EP02688.
XX
PR   02-OCT-1992;   92CH-0003124.
PR   15-JUL-1993;   93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match           100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity  100.0%;  Pred. No. 0.023;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||||||||||||||||||||
Db     200 gcaggtgcatcgccagcatttcgat 224


RESULT   6
AAT68959
ID   AAT68959 standard; DNA; 3084 BP.
XX
AC   AAT68959;
```

XX
DT 05-AUG-1997 (first entry)
XX
DE Candida albicans chitin synthase (CHS1) gene.
XX
KW Chitin synthase; CHS1; yeast-associated disease; antifungal;
KW fungicide; ds.
XX
OS Candida albicans.
XX
PN WO9716540-A1.
XX
PD 09-MAY-1997.
XX
PF 01-NOV-1996; 96WO-US17459.
XX
PR 01-NOV-1995; 95US-0551437.
XX
PA (CHEM-) CHEMGENICS PHARM INC.
XX
PI Bulawa C, Gavrias V, Koltin Y, Riggle P, Winter K;
XX
DR WPI; 1997-272110/24.
DR P-PSDB; AAW17057.
XX
PT New Candida albicans chitin synthase (CHS1) polypeptide and gene -
PT used to develop products for monitoring yeast-associated diseases
PT and for developing antifungal drugs
XX
PS Claim 4; Page 32-36; 64pp; English.
XX
CC A 3084 bp genomic DNA sequence (AAT68959) from Candida albicans
CC codes for a 1027-amino acid chitin synthase (CHS1) (AAW17057), an
CC enzyme essential for cell wall synthesis and yeast cell growth, as
CC well as for colonisation of tissues during infection in vivo.
CC The full-length sequence was isolated using a plasmid rescue
CC strategy and PCR. A novel maltose responsive promoter (see also
CC AAT68960) was also isolated from C. albicans. The isolated CHS1 gene
CC can be incorporated into an expression vector and used to produce
CC CHS1 in host cells. The products can be used for monitoring
CC yeast associated diseases and to develop antifungal drugs (e.g.
CC anti-CHS1 antibodies and CHS1 antisense molecules).
XX
SQ Sequence 3084 BP; 987 A; 550 C; 612 G; 935 T; 0 other;

Query Match 70.4%; Score 17.6; DB 18; Length 3084;
Best Local Similarity 83.3%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 caggtgcatcgccagcatttcgat 25
||||||||| | | |||| |||||
Db 1691 caggtgcattgtctgcatatcgat 1714

RESULT 7
AAT37079
ID AAT37079 standard; cDNA; 3617 BP.
XX
AC AAT37079;
XX
DT 16-OCT-1996 (first entry)
XX
DE AR301 fragment encoding MEI2 protein kinase PAT1.
XX
KW Transformation; host plant; meiosis; spore formation; yeast; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 452..3106
FT /*tag= a
FT /product= Protein kinase PAT1

XX
PN JP08140686-A.
XX
PD 04-JUN-1996.
XX
PF 25-NOV-1994; 94JP-0315514.
XX
PR 25-NOV-1994; 94JP-0315514.
XX
PA (TOYT) TOYOTA JIDOSHA KK.
XX
DR WPI; 1996-316320/32.
DR P-PSDB; AAW00160.
XX
PT Gene coding for plant protein - useful in meiosis and spore
PT formation
XX
PS Claim 1; Page 7-11; 26pp; Japanese.
XX
CC The sequences given in AAT37079-85 are cDNA sequences which were used
CC to transform host plants. The proteins encoded by these sequences
CC are involved in physical activities in plants such as meiosis, etc.
CC in plant cells. The genes are useful for induction of meiosis and
CC spore formation after introduction into a yeast incapable of meiosis
CC and spore formation.
XX
SQ Sequence 3617 BP; 1100 A; 700 C; 767 G; 1050 T; 0 other;

Query Match 70.4%; Score 17.6; DB 17; Length 3617;
Best Local Similarity 83.3%; Pred. No. 67;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 caggtgcatcgccagcatttcgat 25
|||| |||| |||| ||||| |||
Db 1757 caggagcatggccaccattttgat 1780

RESULT 8
AAH18667/c
ID AAH18667 standard; cDNA; 4674 BP.
XX
AC AAH18667;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:18909.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection

```
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 18909; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 4674 BP; 991 A; 1342 C; 1239 G; 1102 T; 0 other;


  Query Match             69.6%;  Score 17.4;  DB 22;  Length 4674;
  Best Local Similarity   94.7%;  Pred. No. 86;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 ggtgcatcgccagcatttc 22
           ||||||||||||||||| ||
Db    2446 GGTGCATCGCCAGCATCTC 2428


RESULT   9
AAV39842/c
ID   AAV39842 standard; DNA; 3186 BP.
XX
AC   AAV39842;
XX
DT   29-SEP-1998  (first entry)
XX
DE   Pseudomonas fluorescens lemA gene.
XX
KW   Pseudomonas; genetic engineering; biocontrol; plant; pathogenic;
KW   Rhizoctonia; Pythium; antifungal; pyrrolnitrin; crop protection; ss.
XX
OS   Pseudomonas fluorescens.
XX
FH   Key             Location/Qualifiers
FT   RBS             245..251
FT                   /*tag=  a
FT                   /note= "potential ribosome binding site"
FT   CDS             256..3006
FT                   /*tag=  b
FT                   /product= "LemA"
XX
PN   WO9824919-A1.
XX
PD   11-JUN-1998.
XX
PF   05-DEC-1997;   97WO-EP06815.
XX
PR   09-SEP-1997;   97US-0058304.
PR   06-DEC-1996;   96US-0761258.
XX
PA   (NOVS ) NOVARTIS AG.
XX
PI   Gaffney TD,  Hill DS,  Lam ST,  Ligon JM,  Stafford JM;
PI   Torkewitz NR;
XX
DR   WPI; 1998-333337/29.
XX
PT   Genetically modified Pseudomonas strains - useful to protect crop
PT   plants by controlling or inhibiting plant pathogen growth, e.g.
PT   growth of Rhizoctonia species
XX
PS   Example 5; Page 63-66; 85pp; English.
XX
CC   A genetically engineered biocontrol strain of Pseudomonas has been
CC   developed that can control attacks on crop plants by pathogenic fungi,
CC   e.g. Rhizoctonia and Pythium and aggressively compete with indigenous
CC   bacteria and microflora in the plant rhizosphere. The strains can be
CC   included with agronomically acceptable carriers or chemical fungicides
CC   (e.g. metalaxyl compounds) in biocontrol compositions. The strains or
CC   compositions can be applied to a plant/plant part to protect it from a
CC   plant pathogenic fungus, by controlling or inhibiting fungal growth.
CC   They can also be applied to the environment in which a plant pathogenic
CC   fungus will grow (e.g. soil) to similarly control or inhibit pathogen
CC   growth, or to seeds to protect plants developing from the seed from a
CC   plant pathogenic fungus. They are especially effective against
CC   Rhizoctonia and Pythium species which cause damping off in cotton.
CC   Rhizoctonia also infects many other crop species (e.g. beans and wheat),
CC   and no effective chemical fungicides are available. The present sequence
CC   represents the Pseudomonas fluorescens lemA gene used in an example from
CC   the present invention.
XX
SQ   Sequence 3186 BP; 646 A; 1091 C; 961 G; 488 T; 0 other;


  Query Match             68.8%;  Score 17.2;  DB 19;  Length 3186;
  Best Local Similarity   86.4%;  Pred. No. 1e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcg 23
           |||| |||| |||||||| |||
Db    2763 CAGGAGCATGGCCAGCATGTCG 2742


RESULT  10
AAV31654/c
ID   AAV31654 standard; DNA; 3186 BP.
XX
AC   AAV31654;
XX
DT   21-AUG-1998  (first entry)
XX
DE   Nucleotide sequence of the Lem A gene.
XX
KW   Lem A gene; phytopathogenic fungi; Rhizoctonia; Pythium spp;
KW   pyrrolnitrin; plant rhizospheres; ss.
XX
OS   Pseudomonas fluorescens.
XX
FH   Key             Location/Qualifiers
FT   CDS             256..3009
FT                   /*tag=  a
FT                   /product= "Lem A protein"
XX
PN   US5756087-A.
XX
PD   26-MAY-1998.
XX
PF   06-DEC-1996;   96US-0761258.
XX
PR   06-DEC-1996;   96US-0761258.
XX
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Gaffney TD,  Hill DS,  Lam ST,  Ligon JM,  Torkewitz N;
```

```
XX
DR   WPI; 1998-321460/28.
XX
PT   Genetically modified Pseudomonas fluorescens strains - useful for
PT   biological control of phytopathogenic fungi
XX
PS   Example 5; Columns 43-48; 35pp; English.
XX
CC   This is the nucleotide sequence of the Pseudomonas fluorescens
CC   Lem a gene, which is used in the method of the invention to produce
CC   modified P. fluorescens strains.  The modified strains are useful for
CC   biological control of phytopathogenic fungi, especially Rhizoctonia
CC   or Pythium spp.  The strains produce high levels of pyrrolnitrin and
CC   are able to compete with indigenous bacteria and microflora in plant
CC   rhizospheres.
XX
SQ   Sequence 3186 BP; 646 A; 1091 C; 961 G; 488 T; 0 other;
```

```
  Query Match             68.8%;  Score 17.2;  DB 19;  Length 3186;
  Best Local Similarity   86.4%;  Pred. No. 1e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcg 23
           |||| |||| |||||||| |||
Db    2763 CAGGAGCATGGCCAGCATGTCG 2742
```

```
RESULT  11
AAX99373/c
ID   AAX99373 standard; DNA; 3186 BP.
XX
AC   AAX99373;
XX
DT   15-NOV-1999  (first entry)
XX
DE   P. fluorescens lemA gene sequence.
XX
KW   Biocontrol strain; Pseudomonas fluorescens; pyrrolnitrin; crop plant;
KW   fungal attack; pathogenic; Rhizoctonia; Phythium; lemA gene; ss.
XX
OS   Pseudomonas fluorescens.
XX
FH   Key             Location/Qualifiers
FT   RBS             245..251
FT                   /*tag=  a
FT                   /note= "potential ribosome binding site"
FT   CDS             256..3006
FT                   /*tag=  b
FT                   /product= "LemA"
XX
PN   US5955348-A.
XX
PD   21-SEP-1999.
XX
PF   25-NOV-1997;   97US-0977306.
XX
PR   25-NOV-1997;   97US-0977306.
XX
PA   (NOVS ) NOVARTIS AG.
XX
PI   Gaffney TD,  Hill DS,  Ligon JM,  Stafford JM,  Torkewitz NR;
XX
DR   WPI; 1999-539577/45.
XX
PT   Strains of Pseudomonas fluorescens with enhanced pyrrolnitrin
PT   production useful for the biocontrol of plant pathogens
XX
PS   Example 5; Columns 51-54; 33pp; English.
XX
CC   The invention provides novel biocontrol strains of P. fluorescens or
CC   their pyrrolnitrin producing progeny. These strains of Pseudomonas have
CC   been genetically engineered to have enhanced biocontrol properties.
CC   (See AAX99368 for the identification names and NRRL accession numbers
CC   for the different biocontrol strains of P. fluorescens). The strains of
CC   P. fluorescens may be used to control pathogenic attack on crop plants,
CC   especially by fungi such as Rhizoctonia or Phythium. The strains produce
CC   enhanced amounts of pyrrolnitrin and aggressively compete with
CC   indigenous bacteria and microflora existing in the rhizosphere of the
CC   plant. The present sequence represents the lemA gene sequence.
XX
SQ   Sequence 3186 BP; 646 A; 1091 C; 961 G; 488 T; 0 other;
```

```
  Query Match             68.8%;  Score 17.2;  DB 20;  Length 3186;
  Best Local Similarity   86.4%;  Pred. No. 1e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcg 23
           |||| |||| |||||||| |||
Db    2763 CAGGAGCATGGCCAGCATGTCG 2742
```

```
RESULT  12
AAZ53592
ID   AAZ53592 standard; DNA; 669 BP.
XX
AC   AAZ53592;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 277 partial DNA sequence SEQ ID NO:1133.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   P-PSDB; AAY74830.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 7; Page 630; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
```

```
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence 669 BP; 133 A; 149 C; 219 G; 168 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 21;  Length 669;
  Best Local Similarity   80.0%;  Pred. No. 98;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||| ||||||||    ||||||
Db      81 gcaggtacatcgccaagcgttcgat 105


RESULT  13
AAZ53594
ID   AAZ53594 standard; DNA; 759 BP.
XX
AC   AAZ53594;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 277 partial DNA sequence SEQ ID NO:1137.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   P-PSDB; AAY74832.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 7; Page 631; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence 759 BP; 138 A; 173 C; 261 G; 187 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 21;  Length 759;
  Best Local Similarity   80.0%;  Pred. No. 1e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||| ||||||||    ||||||
Db     243 gcaggtacatcgccaagcgttcgat 267


RESULT  14
AAX99364
ID   AAX99364 standard; DNA; 6386 BP.
XX
AC   AAX99364;
XX
DT   09-NOV-1999  (first entry)
XX
DE   Maize shrunken-1 (Sh1) gene sequence.
XX
KW   Gene expression; shrunken-1 gene; Sh1 gene; maize; foreign gene; ds.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   exon            131..182
FT                   /*tag= a
FT                   /number= 1
FT   intron          183..1210
FT                   /*tag= b
FT                   /number= 1
FT   exon            1211..1324
FT                   /*tag= c
FT                   /number= 2
FT   intron          1325..1827
FT                   /*tag= d
FT                   /number= 2
FT   exon            1828..1948
FT                   /*tag= e
FT                   /number= 3
FT   intron          1949..2040
FT                   /*tag= f
FT                   /number= 3
FT   exon            2041..2187
FT                   /*tag= g
FT                   /number= 4
FT   intron          2188..2268
FT                   /*tag= h
FT                   /number= 4
FT   exon            2269..2460
FT                   /*tag= i
FT                   /number= 5
FT   intron          2461..2604
FT                   /*tag= j
FT                   /number= 5
FT   exon            2605..2728
FT                   /*tag= k
FT                   /number= 6
FT   intron          2729..2821
FT                   /*tag= l
FT                   /number= 6
FT   exon            2822..3038
FT                   /*tag= m
FT                   /number= 7
FT   intron          3039..3255
```

```
FT                      /*tag=  n
FT                      /number= 7
FT   exon               3256..3351
FT                      /*tag=  o
FT                      /number= 8
FT   intron             3352..3446
FT                      /*tag=  p
FT                      /number= 8
FT   exon               3447..3620
FT                      /*tag=  q
FT                      /number= 9
FT   intron             3621..3701
FT                      /*tag=  r
FT                      /number= 9
FT   exon               3702..3818
FT                      /*tag=  s
FT                      /number= 10
FT   intron             3819..3911
FT                      /*tag=  t
FT                      /number= 10
FT   exon               3912..4078
FT                      /*tag=  u
FT                      /number= 11
FT   intron             4079..4157
FT                      /*tag=  v
FT                      /number= 11
FT   exon               4158..4381
FT                      /*tag=  w
FT                      /number= 12
FT   intron             4382..4516
FT                      /*tag=  x
FT                      /number= 12
FT   exon               4517..4835
FT                      /*tag=  y
FT                      /number= 13
FT   intron             4836..4767
FT                      /*tag=  z
FT                      /number= 13
FT   exon               4768..5212
FT                      /*tag=  aa
FT                      /number= 14
FT   intron             5213..5371
FT                      /*tag=  ab
FT                      /number= 14
FT   exon               5372..5510
FT                      /*tag=  ac
FT                      /number= 15
FT   intron             5511..5635
FT                      /*tag=  ad
FT                      /number= 15
FT   exon               5636..5917
FT                      /*tag=  ae
FT                      /number= 16
XX
PN   US5955330-A.
XX
PD   21-SEP-1999.
XX
PF   07-JUN-1995;   95US-0483376.
XX
PR   18-MAY-1989;   89US-0353854.
PR   05-FEB-1992;   92US-0830956.
PR   04-AUG-1993;   93US-0102115.
PR   07-APR-1995;   95US-0418540.
PR   07-JUN-1995;   95US-0483376.
XX
PA   (RESE ) RESEARCH CORP TECHNOLOGIES INC.
XX
PI   Clancy MA,  Ferl RJ,  Hannah LC,  Vasil IK,  Vasil V;
XX
DR   WPI; 1999-539573/45.
XX
PT   Increasing the expression of a gene in a plant using an intron
XX
PS   Claim 6; Fig 1A-G; 32pp; English.
XX
CC   The invention relates to a method for increasing the expression rate of a
CC   gene in a plant cell using the first intron of the shrunken-1 (Sh1) locus
CC   of maize. The method comprises (a) inserting an intron into the gene
CC   comprising a 5' and 3' splice site of the Sh1 first intron, where the
CC   intron is inserted between a transcription start site and a translation
CC   start site of the gene to form an intron modified gene; and (b)
CC   introducing the modified gene into a plant cell so that the gene is
CC   expressed. The method is useful for increasing the expression of foreign
CC   genes in plant cells. The present sequence represents the maize Sh1
CC   gene sequence.
XX
SQ   Sequence 6386 BP; 1503 A; 1489 C; 1452 G; 1942 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 20;  Length 6386;
  Best Local Similarity   80.0%;  Pred. No. 1.4e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           | ||   ||||||| ||||||||||
Db    5691 ggagaaccatcgcctgcatttcgat 5715


RESULT  15
AAA81455/c
ID   AAA81455 standard; DNA; 23532 BP.
XX
AC   AAA81455;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_3 SEQ ID NO:3.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 253-260; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
```

```
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 23532 BP; 6919 A; 5186 C; 5000 G; 6427 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 21;  Length 23532;
  Best Local Similarity   80.0%;  Pred. No. 1.7e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||||   ||| |||||| ||
Db    2064 GCAGGTGCACGCCCATCATTTCCAT 2040


RESULT  16
AAS08693/c
ID   AAS08693 standard; DNA; 109519 BP.
XX
AC   AAS08693;
XX
DT   26-SEP-2001  (first entry)
XX
DE   Micromonospora DNA encoding biosynthetic enzymes for Everninomycin.
XX
KW   Everninomicin; antibiotic; bottle-neck gene; orthomicin;
KW   fermentation; ds.
XX
OS   Micromonospora carbonacea var. africana.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (132..1382)
FT                   /*tag= a
FT                   /product= "EvdA"
FT   RBS             complement (1389..1394)
FT                   /*tag= b
FT   CDS             complement (1490..2611)
FT                   /*tag= c
FT                   /product= "EvdB"
FT   RBS             complement (2618..2622)
FT                   /*tag= d
FT   CDS             complement (2622..3860)
FT                   /*tag= e
FT                   /product= "EvdC"
FT   RBS             complement (3867..3870)
FT                   /*tag= f
FT   CDS             4143..5312
FT                   /*tag= g
FT                   /product= "EvdD"
FT   RBS             4134..4138
FT                   /*tag= h
FT   CDS             5309..6235
FT                   /*tag= i
FT                   /product= "EvdE"
FT   CDS             6232..7275
FT                   /*tag= j
FT                   /product= "EvdF"
FT   RBS             6226..6229
FT
FT   CDS             7272..8327
FT                   /*tag= l
FT                   /product= "EvdG"
FT   CDS             8342..9364
FT                   /*tag= m
FT                   /product= "EvdH"
FT   RBS             8333..8336
FT                   /*tag= n
FT   CDS             complement (9463..10224)
FT                   /*tag= o
FT                   /product= "EvdI"
FT   RBS             complement (10232..10235)
FT                   /*tag= p
FT   CDS             10424..11176
FT                   /*tag= q
FT                   /product= "EvdJ"
FT   CDS             12027..12455
FT                   /*tag= r
FT                   /product= "EvdK"
FT                   /partial
FT                   /note= "No start codon"
FT   CDS             complement (12108..13022)
FT                   /*tag= s
FT                   /product= "EvdL"
FT   RBS             complement (13027..13030)
FT                   /*tag= t
FT   CDS             complement (14410..15363)
FT                   /*tag= u
FT                   /product= "EvrA"
FT   RBS             complement (15369..15373)
FT                   /*tag= v
FT   CDS             complement (15380..16414)
FT                   /*tag= w
FT                   /product= "EvrB"
FT   CDS             complement 16419..17873
FT                   /*tag= x
FT                   /product= "EvrC"
FT   CDS             complement (17870..18934)
FT                   /*tag= y
FT                   /product= "EvrD"
FT   CDS             19374..20906
FT                   /*tag= z
FT                   /product= "EvrE"
FT   CDS             21064..22542
FT                   /*tag= aa
FT                   /product= "EvrF"
FT   RBS             21056..22542
FT                   /*tag= ab
FT   CDS             22748..24172
FT                   /*tag= ac
FT                   /product= "EvrG"
FT   RBS             22736..22740
FT                   /*tag= ad
FT   CDS             complement (24177..25223)
FT                   /*tag= ae
FT                   /product= "EvrH"
FT   RBS             complement (25230..25233)
FT                   /*tag= af
FT   CDS             25550..26626
FT                   /*tag= ag
FT                   /product= "EvrI"
FT   CDS             26685..30479
FT                   /*tag= ah
FT                   /product= "EvrJ"
FT   RBS             26672..26676
FT                   /*tag= ai
FT   CDS             complement (30557..31876)
FT                   /*tag= aj
FT                   /product= "EvrK"
FT   RBS             complement (31885..31888)
FT                   /*tag= ak
FT   CDS             complement (31941..32882)
```

```
FT                   /*tag= al
FT                   /product= "EvrL"
FT   CDS             complement (33167..34405)
FT                   /*tag= am
FT                   /product= "EvrM"
FT   RBS             complement (34414..34418
FT                   /*tag= an
FT   CDS             complement (34449..35210)
FT                   /*tag= ao
FT                   /product= "EvrN"
FT   RBS             complement (35219..35221)
FT                   /*tag= ap
FT   CDS             complement (35294..36238)
FT                   /*tag= aq
FT                   /product= "EvrO"
FT   CDS             complement (36235..36963)
FT                   /*tag= ar
FT                   /product= "EvrP"
FT   CDS             complement (36998..38026)
FT                   /*tag= as
FT                   /product= "EvrQ"
FT   CDS             complement (38072..38566)
FT                   /*tag= at
FT                   /product= "EvrR"
FT   CDS             complement (38892..40163)
FT                   /*tag= au
FT                   /product= "EvrS"
FT   CDS             complement (40216..40890)
FT                   /*tag= av
FT                   /product= "EvrT"
FT   RBS             complement (40899..40902)
FT                   /*tag= aw
FT   CDS             complement (40887..41576)
FT                   /*tag= ax
FT                   /product= "EvrU"
FT   CDS             complement (41679..42707)
FT                   /*tag= ay
FT                   /product= "EvrV"
FT   RBS             complement (42714..42717)
FT                   /*tag= az
FT   CDS             complement (42810..43799)
FT                   /*tag= ba
FT                   /product= "EvrW"
FT   RBS             complement (43807..43811)
FT                   /*tag= bb
FT   CDS             complement (43799..44866)
FT                   /*tag= bc
FT                   /product= "EvrX"
FT   CDS             complement (45014..45760)
FT                   /*tag= bd
FT                   /product= "EvrY"
FT   RBS             complement (45767..45770)
FT                   /*tag= be
FT   CDS             complement (45962..46714)
FT                   /*tag= bf
FT                   /product= "EvrZ"
FT   RBS             complement (45952..45956)
FT                   /*tag= bg
FT   CDS             complement (47156..49234)
FT                   /*tag= bh
FT                   /product= "EvsA"
FT   CDS             51627..52715
FT                   /*tag= bi
FT                   /product= "EvsB"
FT   RBS             51629..51622
FT                   /*tag= bj
FT   CDS             52889..53557
FT                   /*tag= bk
FT                   /product= "EvsC"
FT   CDS             53554..54207
FT                   /*tag= bl
FT                   /product= "EvbA"
FT   CDS             complement (54362..55117)
FT                   /*tag= bm
FT                   /product= "EvbB"
FT   RBS             complement (55125..55128)
FT                   /*tag= bn
FT   CDS             complement (55135..56094)
FT                   /*tag= bo
FT                   /product= "EvbC"
FT   RBS             complement (56100..56103)
FT                   /*tag= bp
FT   CDS             complement (56184..56813)
FT                   /*tag= bq
FT                   /product= "EvbC2"
FT   CDS             56961..58709
```

```
  Query Match             68.0%;  Score 17;  DB 22;  Length 109519;
  Best Local Similarity   80.0%;  Pred. No. 2.2e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           ||||||||||| ||||| ||||
Db   19908 GCAGGTGCATCATCAGCACCACGAT 19884
```

```
RESULT  17
AAF21610/c
ID   AAF21610 standard; DNA; 349980 BP.
XX
AC   AAF21610;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Neisseria meningitidis B nucleotide sequence SEQ ID NO:111.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; immunogenic; vaccine;
KW   diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW   ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200066791-A1.
XX
PD   09-NOV-2000.
XX
PF   08-MAR-2000; 2000WO-US05928.
XX
PR   30-APR-1999;   99US-0132068.
PR   08-OCT-1999;   99WO-US23573.
PR   28-FEB-2000; 2000GB-0004695.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Pizza M,  Hickey E,  Peterson J,  Tettelin H,  Venter JC,  Masignani V;
PI   Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V,  Rappuoli R;
PI   Frazer CM,  Grandi G;
XX
DR   WPI; 2000-647603/62.
XX
PT   Neisseria meningitidis B full length genome sequence and open reading
PT   frames are used to detect, treat and prevent Neisserial infections -.
XX
PS   Claim 7; Appendix A; 692pp; English.
XX
CC   The present invention describes the full length genome of
CC   Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC   to AAF21613 represent fragments of the NMB genomic sequence, as the
CC   sequence was too long to go in a record on its own it was split into 8
CC   sequences which overlap each other at the beginning and end of each
CC   sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC   the beginning  of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC   the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC   Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC   AAF21606 represent PCR primers which are used in the exemplification of
```

CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX
SQ Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 349980;
Best Local Similarity 80.0%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcaggtgcatcgccagcatttcgat 25
||||||||| ||| |||||| ||
Db 222215 GCAGGTGCACGCCCATCATTTCCAT 222191

RESULT 18
AAA81490/c
ID AAA81490 standard; DNA; 1437668 BP.
XX
AC AAA81490;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;
XX
DR WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 866-1272; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neissariae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

Query Match 68.0%; Score 17; DB 21; Length 1437668;
Best Local Similarity 80.0%; Pred. No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcaggtgcatcgccagcatttcgat 25
||||||||| ||| |||||| ||
Db 1392603 GCAGGTGCACGCCCATCATTTCCAT 1392579

RESULT 19
AAV69543/c
ID AAV69543 standard; DNA; 643 BP.
XX
AC AAV69543;
XX
DT 15-MAR-1999 (first entry)
XX
DE P. neopolydactyla type I polyketide synthase DNA #2.
XX
KW Type I polyketide synthase; soil; lichen; antibiotic biosynthesis;
KW humus; therapeutic; immunosuppressor; antitumour agent; pathogen;
KW genetic diversity; ss.
XX
OS Peltigera neopolydactyla.
XX
PN WO9853097-A2.
XX
PD 26-NOV-1998.
XX
PF 21-MAY-1998; 98WO-CA00488.
XX
PR 22-MAY-1997; 97US-0861774.
XX
PA (TERR-) TERRAGEN DIVERSITY INC.
XX
PI Miao VPW, Seow KT, Waters B, Yap WH;
XX
DR WPI; 1999-070158/06.
DR P-PSDB; AAW82695.
XX
PT New degenerate primers - used for recovering antibiotic biosynthetic
PT DNA from soil/lichen material
XX
PS Claim 17; Page 57; 98pp; English.
XX
CC This sequence encodes a Type I polyketide synthase from the lichen
CC Peltigera neopolydactyla. This protein is used in a method for the
CC recovery of antibiotic biosynthetic DNA from humic materials or lichen.
CC The PCR products of the invention have the potential to be used as
CC therapeutic molecules including antibiotics, immunosuppressors and
CC antitumour agents. The method allows access to the reservoir of genetic
CC diversity in soil pathogenic micro-organisms, in order to find new
CC antibiotics. It also allows access to novel biosynthetic genes/enzymes

```
CC   that can be used to produce antibiotics or produce specific compounds,
CC   enzymatically, in vitro.
XX
SQ   Sequence 643 BP; 175 A; 143 C; 156 G; 169 T; 0 other;


  Query Match             67.2%;  Score 16.8;  DB 20;  Length 643;
  Best Local Similarity   90.0%;  Pred. No. 1.2e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcattt 21
           ||||||| | ||||||||||
Db      53 CAGGTGCTTTGCCAGCATTT 34


RESULT  20
AAT64649
ID   AAT64649 standard; DNA; 65 BP.
XX
AC   AAT64649;
XX
DT   17-JAN-1998  (first entry)
XX
DE   G-CSF receptor agonist gene primer 125for.
XX
KW   Granulocyte colony stimulating factor receptor; agonist; G-CSF;
KW   haematopoietic disorder; neutropenia; bone marrow suppression;
KW   stem cell expansion; gene therapy; circular permutation;
KW   polymerase chain reaction; PCR; primer; ss.
XX
OS   Synthetic.
XX
PN   WO9712977-A1.
XX
PD   10-APR-1997.
XX
PF   04-OCT-1996;   96WO-US15935.
XX
PR   05-OCT-1995;   95US-0004832.
XX
PA   (SEAR ) SEARLE & CO G D.
XX
PI   Braford-Goldberg SR,  Feng Y,  Klein BK,  McKearn JP;
PI   McWherter CA,  Zurfluh LL;
XX
DR   WPI; 1997-244718/22.
XX
PT   Modified human granulocyte colony stimulating factor - useful as
PT   G-CSF receptor agonist for treating haematopoietic disorders, e.g.
PT   neutropenia or bone marrow suppression
XX
PS   Example 25; Page 31; 186pp; English.
XX
CC   This synthetic oligonucleotide comprises primer 125for that was
CC   used in the construction of novel genes (see AAT64612 and AAT64616)
CC   encoding circularly permuted variants (see AAW15040 and AAW15044) of
CC   human granulocyte colony stimulating factor (G-CSF) that act as
CC   G-CSF receptor agonists.  Such receptor agonists can be used in
CC   methods for stimulating production of haematopoietic cells, ex
CC   vivo expansion of stem cells, treatment of haematopoietic disorders
CC   and human gene therapy.
XX
SQ   Sequence 65 BP; 16 A; 19 C; 13 G; 17 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 18;  Length 65;
  Best Local Similarity   82.6%;  Pred. No. 1e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 aggtgcatcgccagcatttcgat 25
           |||||||  ||||||||||  ||
Db      40 aggtgcaatgccagcatttgcat 62


RESULT  21
AAT64616
ID   AAT64616 standard; DNA; 522 BP.
XX
AC   AAT64616;
XX
DT   18-JAN-1998  (first entry)
XX
DE   G-CSF receptor agonist gene in pMON25186.
XX
KW   Granulocyte colony stimulating factor; receptor agonist; G-CSF;
KW   haematopoietic disorder; neutropenia; bone marrow suppression;
KW   stem cell expansion; gene therapy; circular permutation; pMON25186;
KW   ss.
XX
OS   Chimeric Homo sapiens.
OS   Chimeric synthetic.
XX
FH   Key             Location/Qualifiers
FT   mat_peptide     7..519
FT                   /*tag=  a
XX
PN   WO9712977-A1.
XX
PD   10-APR-1997.
XX
PF   04-OCT-1996;   96WO-US15935.
XX
PR   05-OCT-1995;   95US-0004832.
XX
PA   (SEAR ) SEARLE & CO G D.
XX
PI   Braford-Goldberg SR,  Feng Y,  Klein BK,  McKearn JP;
PI   McWherter CA,  Zurfluh LL;
XX
DR   WPI; 1997-244718/22.
DR   P-PSDB; AAW15044.
XX
PT   Modified human granulocyte colony stimulating factor - useful as
PT   G-CSF receptor agonist for treating haematopoietic disorders, e.g.
PT   neutropenia or bone marrow suppression
XX
PS   Claim 7; Page 171; 186pp; English.
XX
CC   This DNA sequence comprises a new N-terminus/C-terminus gene in
CC   pMON25186 that codes for a specific example of a granulocyte colony
CC   stimulating factor (G-CSF) receptor agonist (see AAW15044) comprising
CC   circularly permuted human G-CSF in which a C-terminal fragment of
CC   G-CSF is linked to a truncated N-terminal fragment via a peptide
CC   linker, the breakpoint being at residues 125/126.  It was
CC   constructed using an overlapping PCR primer approach (see AAT64649-
CC   50), resulting in 19 codon changes at the 5' end of the gene.  The
CC   receptor agonist can be expressed in transformed or transfected
CC   host cells (e.g. E. coli).  Claimed receptor agonists (see AAW15034-
CC   38) can be used in claimed methods for stimulating production of
CC   haematopoietic cells in a patient, for selective ex vivo expansion
CC   of stem cells, for treatment of a patient having a haematopoietic
CC   disorder, and in methods of human gene therapy.
XX
SQ   Sequence 522 BP; 94 A; 168 C; 148 G; 112 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 18;  Length 522;
  Best Local Similarity   82.6%;  Pred. No. 1.4e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 aggtgcatcgccagcatttcgat 25
           |||||||  ||||||||||  ||
Db      33 aggtgcaatgccagcatttgcat 55


RESULT  22
```

```
AAT64612
ID   AAT64612 standard; DNA; 531 BP.
XX
AC   AAT64612;
XX
DT   18-JAN-1998  (first entry)
XX
DE   G-CSF receptor agonist gene in pMON25182.
XX
KW   Granulocyte colony stimulating factor; receptor agonist; G-CSF;
KW   haematopoietic disorder; neutropenia; bone marrow suppression;
KW   stem cell expansion; gene therapy; circular permutation; pMON25182;
KW   ss.
XX
OS   Chimeric Homo sapiens.
OS   Chimeric synthetic.
XX
FH   Key             Location/Qualifiers
FT   mat_peptide     7..528
FT                   /*tag=  a
XX
PN   WO9712977-A1.
XX
PD   10-APR-1997.
XX
PF   04-OCT-1996;    96WO-US15935.
XX
PR   05-OCT-1995;    95US-0004832.
XX
PA   (SEAR ) SEARLE & CO G D.
XX
PI   Braford-Goldberg SR,  Feng Y,  Klein BK,  McKearn JP;
PI   McWherter CA,  Zurfluh LL;
XX
DR   WPI; 1997-244718/22.
DR   P-PSDB; AAW15040.
XX
PT   Modified human granulocyte colony stimulating factor - useful as
PT   G-CSF receptor agonist for treating haematopoietic disorders, e.g.
PT   neutropenia or bone marrow suppression
XX
PS   Claim 7; Page 170; 186pp; English.
XX
CC   This DNA sequence comprises a new N-terminus/C-terminus gene in
CC   pMON25182 that codes for a specific example of a granulocyte colony
CC   stimulating factor (G-CSF) receptor agonist (see AAW15040) comprising
CC   circularly permuted human G-CSF in which the N-terminus corresponds
CC   to amino acid residue 126 of a generic G-CSF receptor agonist (see
CC   AAW15025) and the C-terminus to residue 125.  It was constructed
CC   using an overlapping PCR primer method (see also AAT64649-50),
CC   resulting in 19 codon changes at the 5' end of the gene.  The
CC   receptor agonist can be expressed in transformed or transfected
CC   host cells (e.g. E. coli).  Claimed receptor agonists (see AAW15034-
CC   38) can be used in claimed methods for stimulating production of
CC   haematopoietic cells in a patient, for selective ex vivo expansion
CC   of stem cells, for treatment of a patient having a haematopoietic
CC   disorder, and in methods of human gene therapy.
XX
SQ   Sequence 531 BP; 98 A; 176 C; 146 G; 111 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 18;  Length 531;
  Best Local Similarity   82.6%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 aggtgcatcgccagcatttcgat 25
           |||||||  ||||||||||  ||
Db      33 aggtgcaatgccagcatttgcat 55


RESULT  23
AAI42059/c
ID   AAI42059 standard; DNA; 584 BP.
XX
AC   AAI42059;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #10745 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 10745; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 584 BP; 135 A; 188 C; 130 G; 131 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 22;  Length 584;
  Best Local Similarity   82.6%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcg 23
           |||||||  |||  |||||  ||||
Db      74 GCAGGTGGATCACCAGCTCTTCG 52


RESULT  24
AAS22818
ID   AAS22818 standard; cDNA; 1066 BP.
XX
AC   AAS22818;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human cDNA encoding a novel human protein #384.
XX
KW   Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW   immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
KW   anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW   antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW   thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW   Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW   tissue regeneration; immune disorder.
XX
```

```
OS   Homo sapiens.
XX
PN   WO200155437-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02623.
XX
PR   25-JAN-2000; 2000US-0491404.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-451939/48.
DR   P-PSDB; AAU14513.
XX
PT   Isolated polypeptides useful for treating anti-inflammatory diseases,
PT   nervous system disorders, and for regenerating bone and cartilage -
XX
PS   Claim 1; Page 742-743; 894pp; English.
XX
CC   The invention relates to polynucleotides encoding novel human
CC   proteins or their active domains. The polypeptides, polynucleotides and
CC   antibodies raised against the polypeptides are used in a method of
CC   treatment of a mammal and prevention of disorders caused by the aberrant
CC   protein expression or activity.  The polypeptides can be used as
CC   molecular weight markers, food supplements, and in antibody production.
CC   The polypeptides are used to identify compounds which bind to the
CC   polypeptides. Polynucleotides of the invention are used as probes and
CC   primers, for sequencing, for chromosome or gene mapping, in the
CC   production of recombinant proteins, and in generating anti-sense DNA or
CC   RNA and in gene therapy.  Polypeptides of the invention can be used to
CC   target drugs to a tumour, in assays to determine biological activity, to
CC   raise antibodies/elicit an immune response, to determine quantitative
CC   protein levels, as tissue markers, and to isolate receptors or ligands.
CC   Polypeptides of the invention may also be useful in treating platelet
CC   disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC   ligament and/or nerve tissue, wound healing, treating burns, promoting
CC   the proliferation, differentiation and survival of stem cells, as a
CC   contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC   Alzheimer's, Parkinson's and Huntington's diseases, amylotrophic lateral
CC   sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC   fungal infection or from autoimmunity, cancer, allergy, asthma,
CC   graft-versus-host disease, eczema, haemophilia, thrombosis,
CC   anti-inflammatory diseases, nervous system disorders, and infection.
CC   The present sequence encodes a protein of the invention.
XX
SQ   Sequence 1066 BP; 239 A; 269 C; 299 G; 259 T; 0 other;

  Query Match             66.4%;  Score 16.6;  DB 22;  Length 1066;
  Best Local Similarity   82.6%;  Pred. No. 1.6e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcg 23
           || | ||||| |||||||||| |
Db     230 gcggctgcatggccagcatttgg 252


RESULT  25
AAS22582
ID   AAS22582 standard; cDNA; 1255 BP.
XX
AC   AAS22582;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human cDNA encoding a novel human protein #148.
XX
KW   Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW   immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
KW   anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW   antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW   thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW   Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW   tissue regeneration; immune disorder.
XX
OS   Homo sapiens.
XX
PN   WO200155437-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02623.
XX
PR   25-JAN-2000; 2000US-0491404.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-451939/48.
DR   P-PSDB; AAU14277.
XX
PT   Isolated polypeptides useful for treating anti-inflammatory diseases,
PT   nervous system disorders, and for regenerating bone and cartilage -
XX
PS   Claim 1; Page 376-377; 894pp; English.
XX
CC   The invention relates to polynucleotides encoding novel human
CC   proteins or their active domains. The polypeptides, polynucleotides and
CC   antibodies raised against the polypeptides are used in a method of
CC   treatment of a mammal and prevention of disorders caused by the aberrant
CC   protein expression or activity.  The polypeptides can be used as
CC   molecular weight markers, food supplements, and in antibody production.
CC   The polypeptides are used to identify compounds which bind to the
CC   polypeptides. Polynucleotides of the invention are used as probes and
CC   primers, for sequencing, for chromosome or gene mapping, in the
CC   production of recombinant proteins, and in generating anti-sense DNA or
CC   RNA and in gene therapy.  Polypeptides of the invention can be used to
CC   target drugs to a tumour, in assays to determine biological activity, to
CC   raise antibodies/elicit an immune response, to determine quantitative
CC   protein levels, as tissue markers, and to isolate receptors or ligands.
CC   Polypeptides of the invention may also be useful in treating platelet
CC   disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC   ligament and/or nerve tissue, wound healing, treating burns, promoting
CC   the proliferation, differentiation and survival of stem cells, as a
CC   contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC   Alzheimer's, Parkinson's and Huntington's diseases, amylotrophic lateral
CC   sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC   fungal infection or from autoimmunity, cancer, allergy, asthma,
CC   graft-versus-host disease, eczema, haemophilia, thrombosis,
CC   anti-inflammatory diseases, nervous system disorders, and infection.
CC   The present sequence encodes a protein of the invention.
XX
SQ   Sequence 1255 BP; 252 A; 358 C; 377 G; 268 T; 0 other;

  Query Match             66.4%;  Score 16.6;  DB 22;  Length 1255;
  Best Local Similarity   82.6%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcg 23
           || | ||||| |||||||||| |
Db     419 gcggctgcatggccagcatttgg 441


RESULT  26
AAX04354
ID   AAX04354 standard; DNA; 2074 BP.
XX
AC   AAX04354;
XX
DT   13-APR-1999  (first entry)
XX
```

DE Human secreted protein gene 44 clone HETFJ05.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO9856804-A1.
XX
PD 17-DEC-1998.
XX
PF 11-JUN-1998; 98WO-US12125.
XX
PR 02-OCT-1997; 97US-0061060.
PR 13-JUN-1997; 97US-0049547.
PR 13-JUN-1997; 97US-0049548.
PR 13-JUN-1997; 97US-0049549.
PR 13-JUN-1997; 97US-0049550.
PR 13-JUN-1997; 97US-0049606.
PR 13-JUN-1997; 97US-0049607.
PR 13-JUN-1997; 97US-0049608.
PR 13-JUN-1997; 97US-0049609.
PR 13-JUN-1997; 97US-0049610.
PR 13-JUN-1997; 97US-0049611.
PR 13-JUN-1997; 97US-0050566.
PR 13-JUN-1997; 97US-0050901.
PR 13-JUN-1997; 97US-0052989.
PR 08-JUL-1997; 97US-0051919.
PR 18-AUG-1997; 97US-0055984.
PR 12-SEP-1997; 97US-0058665.
PR 12-SEP-1997; 97US-0058668.
PR 12-SEP-1997; 97US-0058669.
PR 12-SEP-1997; 97US-0058750.
PR 12-SEP-1997; 97US-0058971.
PR 12-SEP-1997; 97US-0058972.
PR 12-SEP-1997; 97US-0058975.
PR 02-OCT-1997; 97US-0060834.
PR 02-OCT-1997; 97US-0060841.
PR 02-OCT-1997; 97US-0060844.
PR 02-OCT-1997; 97US-0060865.
PR 02-OCT-1997; 97US-0061059.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Ebner R, Ferrie AM, Feng P, Greene JM, Lafleur DW;
PI Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM, Shi Y, Young P;
PI Yu GL;
XX
DR WPI; 1999-080881/07.
DR P-PSDB; AAW78169.
XX
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
PS Claim 1; Page 210-212; 380pp; English.
XX
CC This sequence represents a nucleic acid molecule which encodes a
CC secreted human protein. The gene number, and the clone it is derived
CC from, are detailed in the descriptor line. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. AAX04302) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 86 novel genes and their fragments (nucleic
CC acid sequences: AAX04311-X04410; amino acid sequences AAW78126-W78225)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 86
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAX04311 for described uses).
XX
SQ Sequence 2074 BP; 485 A; 521 C; 608 G; 456 T; 4 other;

Query Match 66.4%; Score 16.6; DB 20; Length 2074;
Best Local Similarity 82.6%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gcaggtgcatcgccagcatttcg 23
|| | ||||| |||||||||| |
Db 470 gcggctgcatggccagcatttgg 492

RESULT 27
AAH64980/c
ID AAH64980 standard; DNA; 2568 BP.
XX
AC AAH64980;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 15.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR P-PSDB; AAG89761.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 8; SEQ ID NO: 15; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 2568 BP; 585 A; 843 C; 685 G; 455 T; 0 other;

```
  Query Match             66.4%;  Score 16.6;  DB 22;  Length 2568;
  Best Local Similarity   82.6%;  Pred. No. 1.9e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcg 23
           |||| ||||| ||||| ||| ||
Db    1309 GCAGCTGCATTGCCAGAATTGCG 1287


RESULT  28
AAT91360/c
ID   AAT91360 standard; DNA; 4372 BP.
XX
AC   AAT91360;
XX
DT   14-APR-1998  (first entry)
XX
DE   Orf virus genomic DNA BamHI F and partial BamHI C fragments.
XX
KW   Parapoxvirus; vector; vaccine; antigen; HIV; herpes simplex virus;
KW   Echinococcus granulosis; Trichostronglylus; Haemonchus;
KW   Ostertagia; Taenia ovis; ds.
XX
OS   Orf virus strain NZ-2.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..8)
FT                   /*tag=  a
FT                   /label= F1L
FT                   /note= "5' end of F1L coding sequence"
FT   promoter        complement (9..49)
FT                   /*tag=  b
FT                   /label= PF1L
FT                   /note= "putative F1L gene promoter"
FT   CDS             complement (2421..2423)
FT                   /*tag=  c
FT                   /label= F2L
FT                   /note= "F2L gene initiation codon"
FT   misc_feature    2437..2451
FT                   /*tag=  d
FT                   /note= "intergenic region suitable for foreign
FT                         gene insertion"
FT   promoter        complement (2424..2463)
FT                   /*tag=  e
FT                   /label= PF2L
FT                   /note= "putative F2L gene promoter"
FT   promoter        2475..2521
FT                   /*tag=  f
FT                   /label= PF3R
FT                   /note= "putative F3R gene promoter"
FT   CDS             2532..3215
FT                   /*tag=  g
FT                   /label= F3R
FT                   /note= "putative F3R gene"
FT   promoter        3183..3237
FT                   /*tag=  h
FT                   /label= PF4R
FT                   /note= "putative F4R gene promoter"
FT   misc_feature    3216..3220
FT                   /*tag=  i
FT                   /note= "intergenic region suitable for foreign
FT                         gene insertion"
FT   CDS             3240..4196
FT                   /*tag=  j
FT                   /label= F4R
FT                   /note= "putative F4R gene"
FT   promoter        4142..4192
FT                   /*tag=  k
FT                   /label= PC1R
FT                   /note= "putative C1R gene promoter"
FT   misc_feature    3573..3597
FT                   /*tag=  l
FT                   /note= "intergenic region suitable for foreign
FT                         gene insertion"
FT   CDS             4193..4371
FT                   /*tag=  m
FT                   /label= C1R
FT                   /note= "putative C1R gene"
XX
PN   WO9737031-A1.
XX
PD   09-OCT-1997.
XX
PF   27-MAR-1997;   97WO-NZ00040.
XX
PR   29-MAR-1996;   96NZ-0286284.
XX
PA   (UYOT-) UNIV OTAGO.
XX
PI   Lyttle DJ,  Robinson AJ;
XX
DR   WPI; 1997-503116/46.
XX
PT   Parapoxvirus vectors containing exogenous DNA - comprise parapox
PT   virus, especially orf virus, and exogenous DNA encoding a
PT   polypeptide, e.g. an antigen useful in vaccines
XX
PS   Claim 12; Fig 7; 73pp; English.
XX
CC   This genomic DNA molecule comprises the BamHI F fragment and part
CC   of the BamHI C fragment of the orf virus strain NZ-2 genome
CC   incorporating putative promoter sequences for the F1L, F2L, F4R
CC   and C1R genes as well as intergenic regions suitable as foreign
CC   gene insertion sites.  A claimed parapoxvirus vector comprises a
CC   parapox virus (especially an orf virus) containing exogenous DNA.
CC   Vector fragments or variants with equivalent immunological activity
CC   and host cells (especially bovine or ovine testis cells)
CC   incorporating the vector are also claimed.  The vector and
CC   fragments/variants are useful in vaccine production (claimed),
CC   since they can express antigens, useful in medical and veterinary
CC   applications e.g. parasitic disease control in livestock.  Particular
CC   antigens include HIV envelope protein, herpes simplex glycoprotein,
CC   antigens of Echinococcus granulosis, Trichostronglylus, Haemonchus,
CC   Ostertagia and especially Taenia ovis 45W, 16 kd, 18 kd antigens.
CC   The foreign gene is preferably incorporated in a non-essential
CC   region of the virus genome (see AAT91357-61) under control of a
CC   poxvirus promoter, especially orf virus promoter E1L, F1L or F3R
CC   (see AAT91362-64).  The orf virus vectors limit the cross-species
CC   infection common with e.g. vaccinia virus vectors, since the virus
CC   has a narrow host range; orf is also less virulent than vaccinia in
CC   man.
XX
SQ   Sequence 4372 BP; 817 A; 1346 C; 1439 G; 770 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 18;  Length 4372;
  Best Local Similarity   82.6%;  Pred. No. 2e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcga 24
           |||||||| |||| ||||  |||
Db    3257 CAGGTGCAGCGCCCGCATACCGA 3235


RESULT  29
AAX13070/c
ID   AAX13070 standard; DNA; 10391 BP.
XX
AC   AAX13070;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:133.
XX
```

```
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Dillon PJ,  Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 799-804; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer-based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 10391 BP; 3471 A; 1709 C; 2240 G; 2962 T; 9 other;
```

```
  Query Match             66.4%;  Score 16.6;  DB 20;  Length 10391;
  Best Local Similarity   82.6%;  Pred. No. 2.3e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcga 24
           |||||  ||||||| |||||| |
Db    4469 CAGGTTGATCGCCATCATTTCTA 4447
```

```
RESULT  30
AAQ46806/c
ID   AAQ46806 standard; DNA; 29879 BP.
XX
AC   AAQ46806;
XX
DT   22-DEC-1993  (first entry)
XX
DE   eryA region of S. erythraea chromosome.
XX
KW   Saccarapolyspora erythraea; eryA; biosynthesis; polyketide; module;
KW   erythromycin; condensation; elongation; acyl chain growth;
KW   gene replacement; ss.
XX
OS   Saccarapolyspora erythraea.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..10199
FT                   /*tag=  a
FT                   /note= "ORF 1"
FT   CDS             10218..20921
FT                   /*tag=  b
FT                   /note= "ORF 2"
FT   CDS             20922..29879
FT                   /*tag=  c
FT                   /note= "ORF 3"
XX
PN   WO9313663-A.
XX
PD   22-JUL-1993.
XX
PF   17-JAN-1992;   92WO-US00427.
XX
PR   17-JAN-1992;   92WO-US00427.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Donadio S,  Katz L,  McAlpine JB;
XX
DR   WPI; 1993-242804/30.
DR   P-PSDB; AAR44430-32.
XX
PT   Biosynthesis of specific polyketide analogues esp. erythromycin
PT   cpds. - by introducing altered biosynthetic gene-contg. DNA into
PT   microorganisms
XX
PS   Claim 27; Fig 2; 133pp; English.
XX
CC   This sequence represents a fragment of the Saccarapolyspora erythraea
CC   genome, designated eryA.  The polypeptides encoded by this region
CC   are involved in the biosynthesis of the polyketide segment of
CC   erythromycin.  eryA is organised in modules and each module takes care
CC   of one condensation step.  The precise succession of elongation steps
CC   is dictated by the genetic order of the modules.  This fragment may be
CC   specifically altered such that novel polyketide molecules of desired
CC   structure are produced.  Three types of alteration may be produced;
CC   those inactivating a single function in a module which does not arrest
CC   acyl chain growth; those inactivating a single function in a module
CC   which does affect chain growth; and those affecting an entire module.
CC   The mutations may be introduced by gene replacement.
XX
SQ   Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 other;
```

```
  Query Match             66.4%;  Score 16.6;  DB 14;  Length 29879;
  Best Local Similarity   82.6%;  Pred. No. 2.8e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcg 23
           || ||| |||||||||  ||||
Db   14949 GCGGGTCCATCGCCAGCGCTTCG 14927
```

```
RESULT  31
AAH64966/c
ID   AAH64966 standard; DNA; 349980 BP.
XX
AC   AAH64966;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 1.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
```

```
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI   Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR   WPI; 2001-376931/40.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT   -
XX
PS   Claim 7; SEQ ID NO: 1; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence 349980 BP; 79703 A; 91547 C; 98381 G; 80349 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 22;  Length 349980;
  Best Local Similarity   82.6%;  Pred. No. 3.8e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcg 23
           |||| ||||| ||||| ||| ||
Db   13139 GCAGCTGCATTGCCAGAATTGCG 13117


RESULT  32
AAF26295/c
ID   AAF26295 standard; DNA; 2048 BP.
XX
AC   AAF26295;
XX
DT   02-MAY-2001  (first entry)
XX
DE   Pseudomonas sp Type II (xcp) secretion system DNA ORF04962a.
XX
KW   Type II secretion system; xcp secretion system; ORF04962a; ds.
XX
OS   Pseudomonas sp.
XX
PN   WO200107622-A2.
XX
PD   01-FEB-2001.
XX
PF   25-JUL-2000; 2000WO-EP07112.
XX
PR   27-JUL-1999;   99DE-1035086.
XX
PA   (TIGR-) TIGR INST GENOMIC RES.
PA   (QUIA-) QUIAGEN GMBH.
PA   (GBFB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA   (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA   (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
PI   Fraser CM,  Tuemmler B,  Hoheisel J,  Duesterhoeft A,  Hilbert H;
PI   Timmis KN,  Moore E,  Straetz M,  Heim S,  Nelson KE,  Hickey E;
PI   Peterson J;
XX
DR   WPI; 2001-159723/16.
XX
PT   New DNA encoding type II (xcp) secretion systems of the bacteria of
PT   the genus Pseudomonas  -
XX
PS   Claim 3a; Page 24-25; 30pp; German.
XX
CC   This invention describes novel DNA sequences that encode type II (xcp)
CC   secretion systems, especially type II (xcp) secretion systems of the
CC   bacteria of the genus Pseudomonas, for example of the Pseudomanas putida
CC   and Pseudomonas fluorescens group.
XX
SQ   Sequence 2048 BP; 348 A; 712 C; 674 G; 314 T; 0 other;


  Query Match             65.6%;  Score 16.4;  DB 22;  Length 2048;
  Best Local Similarity   94.4%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagca 18
           ||||||||| ||||||||
Db   1593 GCAGGTGCAACGCCAGCA 1576


RESULT  33
AAZ54339
ID   AAZ54339 standard; DNA; 645 BP.
XX
AC   AAZ54339;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 767 partial DNA sequence SEQ ID NO:2627.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   P-PSDB; AAY75577.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
```

XX
PS Claim 7; Page 1246; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 645 BP; 170 A; 138 C; 171 G; 166 T; 0 other;

Query Match 64.8%; Score 16.2; DB 21; Length 645;
Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 gtgcatcgccagcatttcgat 25
|| ||| |||| |||||||||
Db 160 gtacattgccatcatttcgat 180

RESULT 34
AAZ54340
ID AAZ54340 standard; DNA; 645 BP.
XX
AC AAZ54340;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 767 partial DNA sequence SEQ ID NO:2629.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR P-PSDB; AAY75578.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX

PS Claim 7; Page 1247; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 645 BP; 168 A; 147 C; 170 G; 160 T; 0 other;

Query Match 64.8%; Score 16.2; DB 21; Length 645;
Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 gtgcatcgccagcatttcgat 25
|| ||| |||| |||||||||
Db 160 gtacattgccatcatttcgat 180

RESULT 35
AAZ54341
ID AAZ54341 standard; DNA; 645 BP.
XX
AC AAZ54341;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 767 partial DNA sequence SEQ ID NO:2631.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR P-PSDB; AAY75579.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 7; Page 1247-1248; 1453pp; English.

```
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence 645 BP; 170 A; 146 C; 170 G; 159 T; 0 other;

  Query Match             64.8%;  Score 16.2;  DB 21;  Length 645;
  Best Local Similarity   85.7%;  Pred. No. 2.3e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 gtgcatcgccagcatttcgat 25
           || ||| |||| |||||||||
Db     160 gtacattgccatcatttcgat 180


RESULT  36
AAZ53354/c
ID   AAZ53354 standard; DNA; 900 BP.
XX
AC   AAZ53354;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 158 partial DNA sequence SEQ ID NO:657.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   P-PSDB; AAY74592.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 7; Page 452; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence 900 BP; 226 A; 231 C; 232 G; 211 T; 0 other;

  Query Match             64.8%;  Score 16.2;  DB 21;  Length 900;
  Best Local Similarity   85.7%;  Pred. No. 2.4e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 gtgcatcgccagcatttcgat 25
           ||||| |||||||||||| |
Db     264 GTGCACTGCCAGCATTTCGGT 244


RESULT  37
AAZ53355/c
ID   AAZ53355 standard; DNA; 900 BP.
XX
AC   AAZ53355;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 158 partial DNA sequence SEQ ID NO:659.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   P-PSDB; AAY74593.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 7; Page 453; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
```

CC represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 900 BP; 225 A; 234 C; 232 G; 209 T; 0 other;

Query Match 64.8%; Score 16.2; DB 21; Length 900;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 gtgcatcgccagcatttcgat 25
||||| |||||||||||| |
Db 264 GTGCACTGCCAGCATTTCGGT 244

RESULT 38
AAZ53353/c
ID AAZ53353 standard; DNA; 912 BP.
XX
AC AAZ53353;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrheae ORF 158 partial DNA sequence SEQ ID NO:655.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria gonorrheae.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR P-PSDB; AAY74591.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 7; Page 451-452; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 912 BP; 248 A; 229 C; 222 G; 213 T; 0 other;

Query Match 64.8%; Score 16.2; DB 21; Length 912;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 gtgcatcgccagcatttcgat 25
||||| |||||||||||| |
Db 264 GTGCACTGCCAGCATTTCGGT 244

RESULT 39
AAX77011/c
ID AAX77011 standard; DNA; 1896 BP.
XX
AC AAX77011;
XX
DT 10-AUG-1999 (first entry)
XX
DE Partial Sel-1L coding sequence.
XX
KW Sel-1L; Sel-1 like protein; Hip-1; Alzheimer's disease; diabetes; cancer;
KW insulin dependent diabetes mellitus; pancreatic cancer; stroke; therapy;
KW vascular dementia; Parkinson's disease; coronary heart disease;
KW fat metabolism; cholesterol metabolism; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT misc_feature 829..830
FT /*tag= a
FT /note= "between these two nucleotides there is a gap of
FT about 36 nucleotides"
FT misc_feature 1073..1074
FT /*tag= b
FT /note= "between these two nucleotides there is a gap of
FT about 180 nucleotides"
XX
PN WO9927088-A2.
XX
PD 03-JUN-1999.
XX
PF 19-NOV-1998; 98WO-CA01058.
XX
PR 28-JUL-1998; 98US-0123549.
PR 19-NOV-1997; 97US-0066140.
XX
PA (MOUN) MOUNT SINAI HOSPITAL.
XX
PI Bernstein A, Donoviel D;
XX
DR WPI; 1999-357833/30.
DR P-PSDB; AAY18094.
XX
PT New Sel-1L nucleic acid molecule useful in the treatment of
PT Alzheimer's disease, diabetes and cancer
XX
PS Claim 2; Page 63-64; 77pp; English.
XX
CC This sequence encodes a Sel-1L (Sel-1 like) protein (also previously

```
CC   known as Hip-1) of the invention. A vector containing the Sel-1L DNA, and
CC   the host cell containing it can be used to prepare a Sel-1L protein.
CC   Compositions containing the Sel-1L proteins, or peptides that interfere
CC   with their binding can be used in a method for treating or preventing
CC   Alzheimer's disease, diabetes (especially insulin dependent diabetes
CC   mellitus), cancer (especially pancreatic cancer), stroke, vascular
CC   dementia, Parkinson's disease, or coronary heart disease. The
CC   compositions can also be used to treat conditions requiring modulation of
CC   fat or cholesterol metabolism.
XX
SQ   Sequence 1896 BP; 451 A; 450 C; 479 G; 464 T; 52 other;

  Query Match             64.8%;  Score 16.2;  DB 20;  Length 1896;
  Best Local Similarity   85.7%;  Pred. No. 2.8e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcattt 21
           ||| ||| || ||||||||||
Db     633 GCATGTGAATTGCCAGCATTT 613


RESULT  40
AAX77014/c
ID   AAX77014 standard; DNA; 2109 BP.
XX
AC   AAX77014;
XX
DT   10-AUG-1999  (first entry)
XX
DE   Human Sel-1L coding sequence.
XX
KW   Sel-1L; Sel-1 like protein; Hip-1; Alzheimer's disease; diabetes; cancer;
KW   insulin dependent diabetes mellitus; pancreatic cancer; stroke; therapy;
KW   vascular dementia; Parkinson's disease; coronary heart disease;
KW   fat metabolism; cholesterol metabolism; ss.
XX
OS   Homo sapiens.
XX
PN   WO9927088-A2.
XX
PD   03-JUN-1999.
XX
PF   19-NOV-1998;   98WO-CA01058.
XX
PR   28-JUL-1998;   98US-0123549.
PR   19-NOV-1997;   97US-0066140.
XX
PA   (MOUN ) MOUNT SINAI HOSPITAL.
XX
PI   Bernstein A,  Donoviel D;
XX
DR   WPI; 1999-357833/30.
DR   P-PSDB; AAY18097.
XX
PT   New Sel-1L nucleic acid molecule useful in the treatment of
PT   Alzheimer's disease, diabetes and cancer
XX
PS   Claim 2; Page 75-76; 77pp; English.
XX
CC   This sequence encodes a Sel-1L (Sel-1 like) protein (also previously
CC   known as Hip-1) of the invention. A vector containing the Sel-1L DNA, and
CC   the host cell containing it can be  ed to prepare a Sel-1L protein.
CC   Compositions containing the Sel-1L proteins, or peptides that interfere
CC   with their binding can be used in a method for treating or preventing
CC   Alzheimer's disease, diabetes (especially insulin dependent diabetes
CC   mellitus), cancer (especially pancreatic cancer), stroke, vascular
CC   dementia, Parkinson's disease, or coronary heart disease. The
CC   compositions can also be used to treat conditions requiring modulation of
CC   fat or cholesterol metabolism.
XX
SQ   Sequence 2109 BP; 591 A; 442 C; 492 G; 584 T; 0 other;

  Query Match             64.8%;  Score 16.2;  DB 20;  Length 2109;
  Best Local Similarity   85.7%;  Pred. No. 2.8e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcattt 21
           ||| ||| || ||||||||||
Db     136 GCATGTGAATTGCCAGCATTT 116


RESULT  41
AAF93859/c
ID   AAF93859 standard; cDNA; 2144 BP.
XX
AC   AAF93859;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Human cDNA encoding a membrane or secretory protein clone PSEC0209.
XX
KW   Human; secretory protein; membrane protein; vaccine; gene therapy;
KW   rheumatoid arthritis; diabetes; ss.
XX
OS   Homo sapiens.
XX
PN   EP1067182-A2.
XX
PD   10-JAN-2001.
XX
PF   07-JUL-2000; 2000EP-0114090.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   11-JAN-2000; 2000JP-0118775.
PR   02-MAY-2000; 2000JP-0183766.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T, . Nishikawa T,  Kawai Y,  Sugiyama T,  Hayashi K;
XX
DR   WPI; 2001-093989/11.
DR   P-PSDB; AAB88432.
XX
PT   Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT   gene therapy or as candidate target molecules in drug development -
XX
PS   Claim 1; SEQ ID 231; 609pp + CD ROM; English.
XX
CC   This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC   which encode human secretory or membrane proteins represented by
CC   AAB88317 - AAB88419. Included in the invention are primers
CC   AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC   cDNA sequences of the invention. The invention also includes methods for
CC   the production of antibodies directed against the proteins, and cDNA
CC   sequences, which can be used in vaccines. The polynucleotide sequences
CC   can be used in gene therapy. The polynucleotide sequences and the
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with inappropriate secretory
CC   protein/membrane protein expression. The nucleic acids and complementary
CC   sequences may also be used as DNA probes in diagnostic assays
CC   (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC   presence of similar nucleic acid sequences in samples. They may also be
CC   used to study the expression and function of secretory proteins/membrane
CC   polypeptides and their role in metabolism. The polypeptides may be used
CC   as antigens in the production of antibodies against them and in assays to
CC   identify modulators (agonists and antagonists) of expression and
CC   activity. The antibodies and antagonists may also be used as therapeutic
CC   agents to down regulate expression and activity. The antibodies may also
CC   be used as diagnostic agents for detecting the presence of the
CC   polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC   (ELTISA). Examples of diseases which may be treated include rheumatoid
CC   arthritis and diabetes.
XX
SQ   Sequence 2144 BP; 623 A; 407 C; 552 G; 562 T; 0 other;
```

```
  Query Match             64.8%;  Score 16.2;  DB 22;  Length 2144;
  Best Local Similarity   85.7%;  Pred. No. 2.8e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcattt 21
           ||| ||| || ||||||||||
Db    1331 GCATGTGAATTGCCAGCATTT 1311


RESULT  42
AAX76577/c
ID   AAX76577 standard; cDNA; 2382 BP.
XX
AC   AAX76577;
XX
DT   10-AUG-1999  (first entry)
XX
DE   Human pancreas-specific tumour suppressor TSA305 gene SEQ ID NO:2.
XX
KW   Human; pancreas; TSA305; pancreatic cancer; carcinoma; diagnosis;
KW   therapy; tumour suppressor; ss.
XX
OS   Homo sapiens.
XX
PN   WO9928457-A1.
XX
PD   10-JUN-1999.
XX
PF   25-NOV-1998;   98WO-JP05306.
XX
PR   20-APR-1998;   98JP-0126803.
PR   28-NOV-1997;   97JP-0343789.
XX
PA   (SAKA ) OTSUKA PHARM CO LTD.
XX
PI   Harada Y,  Ozaki K;
XX
DR   WPI; 1999-358128/30.
DR   P-PSDB; AAY17750.
XX
PT   Tumour suppressor gene TSA305 expressed specifically in pancreas,
PT   useful for diagnosis and treatment of pancreatic cancer
XX
PS   Claim 2; Page 42-43; 54pp; Japanese.
XX
CC   The present sequence represents a gene designated TSA305, which is
CC   expressed specifically in pancreas tissue. The gene, sequences
CC   hybridizing with it, its expression product, and antibodies recognizing
CC   the expression product, are useful in the investigation, diagnosis,
CC   prevention and treatment of pancreatic cancer.
XX
SQ   Sequence 2382 BP; 689 A; 501 C; 614 G; 578 T; 0 other;


  Query Match             64.8%;  Score 16.2;  DB 20;  Length 2382;
  Best Local Similarity   85.7%;  Pred. No. 2.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcattt 21
           ||| ||| || ||||||||||
Db    1235 GCATGTGAATTGCCAGCATTT 1215


RESULT  43
AAX77013/c
ID   AAX77013 standard; DNA; 3666 BP.
XX
AC   AAX77013;
XX
DT   10-AUG-1999  (first entry)
XX
DE   Mouse Sel-1L splice variant coding sequence.
XX
KW   Sel-1L; Sel-1 like protein; Hip-1; Alzheimer's disease; diabetes; cancer;
KW   insulin dependent diabetes mellitus; pancreatic cancer; stroke; therapy;
KW   vascular dementia; Parkinson's disease; coronary heart disease;
KW   fat metabolism; cholesterol metabolism; ss.
XX
OS   Mus sp.
XX
PN   WO9927088-A2.
XX
PD   03-JUN-1999.
XX
PF   19-NOV-1998;   98WO-CA01058.
XX
PR   28-JUL-1998;   98US-0123549.
PR   19-NOV-1997;   97US-0066140.
XX
PA   (MOUN ) MOUNT SINAI HOSPITAL.
XX
PI   Bernstein A,  Donoviel D;
XX
DR   WPI; 1999-357833/30.
DR   P-PSDB; AAY18096.
XX
PT   New Sel-1L nucleic acid molecule useful in the treatment of
PT   Alzheimer's disease, diabetes and cancer
XX
PS   Claim 2; Page 71-73; 77pp; English.
XX
CC   This sequence encodes a Sel-1L (Sel-1 like) protein (also previously
CC   known as Hip-1) of the invention. A vector containing the Sel-1L DNA, and
CC   the host cell containing it can be used to prepare a Sel-1L protein.
CC   Compositions containing the Sel-1L proteins, or peptides that interfere
CC   with their binding can be used in a method for treating or preventing
CC   Alzheimer's disease, diabetes (especially insulin dependent diabetes
CC   mellitus), cancer (especially pancreatic cancer), stroke, vascular
CC   dementia, Parkinson's disease, or coronary heart disease. The
CC   compositions can also be used to treat conditions requiring modulation of
CC   fat or cholesterol metabolism.
XX
SQ   Sequence 3666 BP; 956 A; 881 C; 953 G; 874 T; 2 other;


  Query Match             64.8%;  Score 16.2;  DB 20;  Length 3666;
  Best Local Similarity   85.7%;  Pred. No. 3.1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcattt 21
           ||| ||| || ||||||||||
Db    1148 GCATGTGAATTGCCAGCATTT 1128


RESULT  44
AAX77012/c
ID   AAX77012 standard; DNA; 3667 BP.
XX
AC   AAX77012;
XX
DT   10-AUG-1999  (first entry)
XX
DE   Full length Mouse Sel-1L coding sequence.
XX
KW   Sel-1L; Sel-1 like protein; Hip-1; Alzheimer's disease; diabetes; cancer;
KW   insulin dependent diabetes mellitus; pancreatic cancer; stroke; therapy;
KW   vascular dementia; Parkinson's disease; coronary heart disease;
KW   fat metabolism; cholesterol metabolism; ss.
XX
OS   Mus sp.
XX
PN   WO9927088-A2.
XX
PD   03-JUN-1999.
```

XX
PF 19-NOV-1998; 98WO-CA01058.
XX
PR 28-JUL-1998; 98US-0123549.
PR 19-NOV-1997; 97US-0066140.
XX
PA (MOUN) MOUNT SINAI HOSPITAL.
XX
PI Bernstein A, Donoviel D;
XX
DR WPI; 1999-357833/30.
DR P-PSDB; AAY18095.
XX
PT New Sel-1L nucleic acid molecule useful in the treatment of
PT Alzheimer's disease, diabetes and cancer
XX
PS Claim 2; Page 66-69; 77pp; English.
XX
CC This sequence encodes a Sel-1L (Sel-1 like) protein (also previously
CC known as Hip-1) of the invention. A vector containing the Sel-1L DNA, and
CC the host cell containing it can be used to prepare a Sel-1L protein.
CC Compositions containing the Sel-1L proteins, or peptides that interfere
CC with their binding can be used in a method for treating or preventing
CC Alzheimer's disease, diabetes (especially insulin dependent diabetes
CC mellitus), cancer (especially pancreatic cancer), stroke, vascular
CC dementia, Parkinson's disease, or coronary heart disease. The
CC compositions can also be used to treat conditions requiring modulation of
CC fat or cholesterol metabolism.
XX
SQ Sequence 3667 BP; 956 A; 881 C; 954 G; 874 T; 2 other;

Query Match 64.8%; Score 16.2; DB 20; Length 3667;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gcaggtgcatcgccagcattt 21
||| ||| || ||||||||||
Db 1148 GCATGTGAATTGCCAGCATTT 1128

RESULT 45
AAI61089/c
ID AAI61089 standard; cDNA; 5662 BP.
XX
AC AAI61089;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5078.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM41933.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 5078; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 5662 BP; 1631 A; 1126 C; 1247 G; 1658 T; 0 other;

Query Match 64.8%; Score 16.2; DB 22; Length 5662;
Best Local Similarity 85.7%; Pred. No. 3.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 ggtgcatcgccagcatttcga 24
|| |||||||||||||| |
Db 4776 GGATCATCGCCAGCATTTCAA 4756

Search completed: December 26, 2001, 12:19:02
Job time: 6903 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:23:53 ; Search time 141.92 Seconds
(without alignments)
39.895 Million cell updates/sec

Title: US-09-396-196F-6
Perfect score: 25
Sequence: 1 gcaggtgcatcgccagcatttcgat 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
	2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
	3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
	4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
	5	17.6	70.4	3084	1	US-08-551-437-1	Sequence 1, Appli
	6	17.6	70.4	3084	3	US-09-004-225-1	Sequence 1, Appli
	7	17.6	70.4	3084	4	US-09-084-346-1	Sequence 1, Appli
	8	17.6	70.4	3084	4	US-09-104-704-1	Sequence 1, Appli
c	9	17.2	68.8	3186	1	US-08-761-258-8	Sequence 8, Appli
c	10	17.2	68.8	3186	2	US-08-977-306-8	Sequence 8, Appli
	11	17	68.0	6386	2	US-08-483-376-1	Sequence 1, Appli
c	12	16.8	67.2	643	4	US-08-861-774E-53	Sequence 53, Appl
	13	16.6	66.4	65	3	US-08-833-167-19	Sequence 19, Appl
	14	16.6	66.4	522	3	US-08-833-167-40	Sequence 40, Appl
	15	16.6	66.4	531	3	US-08-833-167-36	Sequence 36, Appl
c	16	16.6	66.4	20235	1	US-07-642-734C-3	Sequence 3, Appli
c	17	16.6	66.4	20235	3	US-08-439-009A-3	Sequence 3, Appli
	18	16.2	64.8	8878	1	US-08-759-444-2	Sequence 2, Appli
	19	16.2	64.8	9880	3	US-08-680-897-1	Sequence 1, Appli
c	20	16	64.0	750	3	US-09-167-717-3	Sequence 3, Appli
c	21	16	64.0	2263	2	US-08-687-865A-1	Sequence 1, Appli
c	22	16	64.0	2263	4	US-09-043-711-1	Sequence 1, Appli
	23	16	64.0	4203	2	US-08-866-757-1	Sequence 1, Appli
	24	16	64.0	4203	4	US-09-153-593-1	Sequence 1, Appli
	25	16	64.0	11236	1	US-07-853-913-1	Sequence 1, Appli
	26	16	64.0	30001	1	US-08-125-468-1	Sequence 1, Appli
	27	16	64.0	30001	2	US-08-474-933-1	Sequence 1, Appli
	28	16	64.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	29	16	64.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	30	15.6	62.4	386	1	US-08-620-467A-9	Sequence 9, Appli
c	31	15.6	62.4	386	1	US-08-348-572-9	Sequence 9, Appli
c	32	15.6	62.4	386	3	US-09-041-090B-9	Sequence 9, Appli
c	33	15.6	62.4	827	4	US-08-998-416-315	Sequence 315, App
c	34	15.6	62.4	1126	1	US-08-233-788A-48	Sequence 48, Appl
c	35	15.6	62.4	3172	6	5196511-1	Patent No. 5196511
	36	15.6	62.4	12732	4	US-09-060-756-1	Sequence 1, Appli
	37	15.4	61.6	480	6	5208144-34	Patent No. 5208144
c	38	15.4	61.6	644	4	US-08-998-416-155	Sequence 155, App
	39	15.4	61.6	1542	4	US-09-385-028-16	Sequence 16, Appl
	40	15.4	61.6	2266	2	US-08-724-394A-18	Sequence 18, Appl
c	41	15.4	61.6	3900	2	US-08-356-060A-42	Sequence 42, Appl
c	42	15.4	61.6	3900	4	US-08-460-900C-42	Sequence 42, Appl
c	43	15.4	61.6	3900	4	US-08-674-509B-42	Sequence 42, Appl
c	44	15.4	61.6	3900	4	US-08-954-698-42	Sequence 42, Appl
c	45	15.4	61.6	4104	4	US-08-881-706-1	Sequence 1, Appli
c	46	15.4	61.6	4434	2	US-08-540-406-5	Sequence 5, Appli
c	47	15.4	61.6	4434	3	US-08-656-055-5	Sequence 5, Appli
c	48	15.4	61.6	4434	4	US-08-954-668-5	Sequence 5, Appli
c	49	15.4	61.6	4434	5	PCT-US95-13233-5	Sequence 5, Appli
	50	15.4	61.6	11604	4	US-09-385-028-13	Sequence 13, Appl
	51	15.4	61.6	15079	4	US-09-385-028-1	Sequence 1, Appli
c	52	15.2	60.8	79	1	US-08-275-487-14	Sequence 14, Appl
	53	15.2	60.8	81	1	US-08-275-487-13	Sequence 13, Appl
	54	15.2	60.8	126	1	US-08-275-488A-8	Sequence 8, Appli
	55	15.2	60.8	126	1	US-08-275-490-8	Sequence 8, Appli
	56	15.2	60.8	126	1	US-08-446-380-8	Sequence 8, Appli
	57	15.2	60.8	126	1	US-08-446-374-8	Sequence 8, Appli
	58	15.2	60.8	126	1	US-08-446-382-8	Sequence 8, Appli
	59	15.2	60.8	126	1	US-08-445-801-8	Sequence 8, Appli
	60	15.2	60.8	126	1	US-08-275-487-8	Sequence 8, Appli
	61	15.2	60.8	126	5	PCT-US95-08919-8	Sequence 8, Appli
c	62	15.2	60.8	954	1	US-08-624-125-9	Sequence 9, Appli
	63	15.2	60.8	1485	1	US-08-275-488A-3	Sequence 3, Appli
	64	15.2	60.8	1485	1	US-08-275-490-3	Sequence 3, Appli
	65	15.2	60.8	1485	1	US-08-446-380-3	Sequence 3, Appli
	66	15.2	60.8	1485	1	US-08-446-374-3	Sequence 3, Appli
	67	15.2	60.8	1485	1	US-08-446-382-3	Sequence 3, Appli
	68	15.2	60.8	1485	1	US-08-445-801-3	Sequence 3, Appli
	69	15.2	60.8	1485	1	US-08-275-487-3	Sequence 3, Appli
	70	15.2	60.8	1485	5	PCT-US95-08919-3	Sequence 3, Appli
	71	15.2	60.8	1620	1	US-08-275-488A-5	Sequence 5, Appli
	72	15.2	60.8	1620	1	US-08-275-490-5	Sequence 5, Appli
	73	15.2	60.8	1620	1	US-08-446-380-5	Sequence 5, Appli
	74	15.2	60.8	1620	1	US-08-446-374-5	Sequence 5, Appli
	75	15.2	60.8	1620	1	US-08-446-382-5	Sequence 5, Appli
	76	15.2	60.8	1620	1	US-08-445-801-5	Sequence 5, Appli
	77	15.2	60.8	1620	1	US-08-275-487-5	Sequence 5, Appli
	78	15.2	60.8	1620	5	PCT-US95-08919-5	Sequence 5, Appli
c	79	15.2	60.8	1956	2	US-08-762-308-10	Sequence 10, Appl
	80	15.2	60.8	3029	1	US-08-275-488A-6	Sequence 6, Appli
	81	15.2	60.8	3029	1	US-08-275-490-6	Sequence 6, Appli
	82	15.2	60.8	3029	1	US-08-446-380-6	Sequence 6, Appli
	83	15.2	60.8	3029	1	US-08-446-374-6	Sequence 6, Appli
	84	15.2	60.8	3029	1	US-08-446-382-6	Sequence 6, Appli
	85	15.2	60.8	3029	1	US-08-445-801-6	Sequence 6, Appli
	86	15.2	60.8	3029	1	US-08-275-487-6	Sequence 6, Appli
	87	15.2	60.8	3029	5	PCT-US95-08919-6	Sequence 6, Appli
c	88	15.2	60.8	6831	2	US-08-609-049A-27	Sequence 27, Appl
c	89	15.2	60.8	6831	4	US-09-170-996-27	Sequence 27, Appl
	90	15.2	60.8	9562	3	US-08-886-967-1	Sequence 1, Appli
	91	15.2	60.8	9562	4	US-09-306-949-1	Sequence 1, Appli
c	92	15.2	60.8	11873	2	US-08-970-269A-32	Sequence 32, Appl
c	93	15.2	60.8	11873	4	US-09-407-562-32	Sequence 32, Appl
c	94	15.2	60.8	11878	2	US-08-970-269A-31	Sequence 31, Appl
c	95	15.2	60.8	11878	4	US-09-407-562-31	Sequence 31, Appl
c	96	15.2	60.8	11883	2	US-08-970-269A-28	Sequence 28, Appl
c	97	15.2	60.8	11883	4	US-09-407-562-28	Sequence 28, Appl
c	98	15	60.0	292	1	US-07-691-191A-5	Sequence 5, Appli
c	99	15	60.0	292	1	US-08-202-486-5	Sequence 5, Appli
c	100	15	60.0	840	4	US-08-998-416-341	Sequence 341, App

ALIGNMENTS

```
RESULT   1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Enhanced Biotin Biosynthesis in Plant Tissue
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/351,970
;      FILING DATE:  08-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-401-068-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.0067;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||||||||||||||||||||
Db      84 GCAGGTGCATCGCCAGCATTTCGAT 108


RESULT   2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-846-338-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.0067;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||||||||||||||||||||
Db      84 GCAGGTGCATCGCCAGCATTTCGAT 108


RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A-15/9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1157
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 117
;      OTHER INFORMATION:  /product= "Biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioB"
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2295..3050
;      OTHER INFORMATION:  /codon_start= 2295
;      OTHER INFORMATION:  /function= "involved in pimeloyl-CoA synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /gene= "bioC"
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3750..5039
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3750
;      OTHER INFORMATION:  /EC_number= 2.6.1.62
;      OTHER INFORMATION:  /product= "DAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioA"
;      OTHER INFORMATION:  /number= 5
;      OTHER INFORMATION:  /standard_name=
;      OTHER INFORMATION:  "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
;      OTHER INFORMATION:  aminotransf."
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  5098..5574
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 5098
;      OTHER INFORMATION:  /function= "unknown, involved in biotin synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "ORFI"
;      OTHER INFORMATION:  /number= 6
;    FEATURE:
;      NAME/KEY:  -10_signal
;      LOCATION:  45..49
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  -35_signal
;      LOCATION:  23..28
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  105..119
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "bioB RBS no.9"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  2284..2297
;      OTHER INFORMATION:  /standard_name= "bioC RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3742..3752
;      OTHER INFORMATION:  /standard_name= "bioA RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  5088..5100
;      OTHER INFORMATION:  /standard_name= "ORFI RBS"
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  5583..5644
;      OTHER INFORMATION:  /standard_name= "rho-independent
;      OTHER INFORMATION:  transcriptional terminator"
;    FEATURE:
;      NAME/KEY:  stem_loop
;      LOCATION:  5583..5605
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..96
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "promoter ptac"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-1


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0088;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||||||||||||||||||||
Db     200 GCAGGTGCATCGCCAGCATTTCGAT 224


RESULT   4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A15-9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1154..2308
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1154
;      OTHER INFORMATION:  /EC_number= 2.3.1.47
;      OTHER INFORMATION:  /product= "KAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "8-Amino-7-oxononanoate synthase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3043..3753
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3043
;      OTHER INFORMATION:  /EC_number= 6.3.3.3
;      OTHER INFORMATION:  /product= "DTB synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioD"
;      OTHER INFORMATION:  /number= 4
;      OTHER INFORMATION:  /standard_name= "Dethiobiotin synthase"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1141..1156
;      OTHER INFORMATION:  /standard_name= "bioF RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3030..3045
;      OTHER INFORMATION:  /standard_name= "bioD RBS"
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-6


  Query Match           100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity  100.0%;  Pred. No. 0.0088;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcgat 25
          |||||||||||||||||||||||||
Db    200 GCAGGTGCATCGCCAGCATTTCGAT 224


RESULT   5
US-08-551-437-1
; Sequence 1, Application US/08551437
; Patent No. 5824545
;  GENERAL INFORMATION:
;    APPLICANT:  Koltin, Yigal
;    APPLICANT:  Riggle, Perry
;    APPLICANT:  Gavrias, Vicky
;    APPLICANT:  Bulawa, Chris
;    APPLICANT:  Winter, Ben
;    TITLE OF INVENTION:  IDENTIFICATION OF EUKARYOTIC
;    TITLE OF INVENTION:  GROWTH-RELATED GENES AND PROMOTER ISOLATION VECTOR
;    TITLE OF INVENTION:  AND METHOD OF USE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/551,437
;      FILING DATE:  01-NOV-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  06286/009001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-507
;      TELEFAX:  617/542-890
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3084 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1...3081
US-08-551-437-1


  Query Match           70.4%;  Score 17.6;  DB 1;  Length 3084;
  Best Local Similarity  83.3%;  Pred. No. 21;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 caggtgcatcgccagcatttcgat 25
          |||||||||  | | ||||  |||||
Db   1691 CAGGTGCATTGTCTGCATATCGAT 1714


RESULT   6
US-09-004-225-1
; Sequence 1, Application US/09004225
; Patent No. 6020133
;  GENERAL INFORMATION:
;    APPLICANT:  Koltin, Yigal
;    APPLICANT:  Riggle, Perry
;    APPLICANT:  Gavrias, Vicky
;    APPLICANT:  Bulawa, Chris
;    APPLICANT:  Winter, Ben
;    TITLE OF INVENTION:  IDENTIFICATION OF EUKARYOTIC
;    TITLE OF INVENTION:  GROWTH-RELATED GENES AND PROMOTER ISOLATION VECTOR
;    TITLE OF INVENTION:  AND METHOD OF USE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
```

```
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  US
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  Windows95
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/004,225
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/551,437
;       FILING DATE:  01-NOV-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fasse, J. Peter
;       REGISTRATION NUMBER:  32,983
;       REFERENCE/DOCKET NUMBER:  06286/009001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-507
;       TELEFAX:  617/542-890
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3084 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  Coding Sequence
;       LOCATION:  1...3081
US-09-004-225-1


  Query Match             70.4%;  Score 17.6;  DB 3;  Length 3084;
  Best Local Similarity   83.3%;  Pred. No. 21;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcgat 25
           ||||||||| | | |||| |||||
Db    1691 CAGGTGCATTGTCTGCATATCGAT 1714


RESULT   7
US-09-084-346-1
; Sequence 1, Application US/09084346
; Patent No. 6251593
;  GENERAL INFORMATION:
;    APPLICANT:  Koltin, Yigal
;    APPLICANT:  Riggle, Perry
;    APPLICANT:  Gavrias, Vicky
;    APPLICANT:  Bulawa, Chris
;    APPLICANT:  Winter, Ben
;    TITLE OF INVENTION:  IDENTIFICATION OF EUKARYOTIC
;    TITLE OF INVENTION:  GROWTH-RELATED GENES AND PROMOTER ISOLATION VECTOR
;    TITLE OF INVENTION:  AND METHOD OF USE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/084,346
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/551,437
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  06286/009001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-507
;      TELEFAX:  617/542-890
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3084 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1...3081
US-09-084-346-1


  Query Match             70.4%;  Score 17.6;  DB 4;  Length 3084;
  Best Local Similarity   83.3%;  Pred. No. 21;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcgat 25
           ||||||||| | | |||| |||||
Db    1691 CAGGTGCATTGTCTGCATATCGAT 1714


RESULT   8
US-09-104-704-1
; Sequence 1, Application US/09104704
; Patent No. 6291218
;  GENERAL INFORMATION:
;    APPLICANT:  Koltin, Yigal
;    APPLICANT:  Riggle, Perry
;    APPLICANT:  Gavrias, Vicky
;    APPLICANT:  Bulawa, Chris
;    APPLICANT:  Winter, Ben
;    TITLE OF INVENTION:  IDENTIFICATION OF EUKARYOTIC
;    TITLE OF INVENTION:  GROWTH-RELATED GENES AND PROMOTER ISOLATION VECTOR
;    TITLE OF INVENTION:  AND METHOD OF USE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/104,704
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/551,437
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fasse, J. Peter
;      REGISTRATION NUMBER:  32,983
;      REFERENCE/DOCKET NUMBER:  06286/009001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-507
;      TELEFAX:  617/542-890
;      TELEX:  200154
```

```
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3084 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1...3081
US-09-104-704-1


  Query Match             70.4%;  Score 17.6;  DB 4;  Length 3084;
  Best Local Similarity   83.3%;  Pred. No. 21;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 caggtgcatcgccagcatttcgat 25
          ||||||||| | | |||| |||||
Db   1691 CAGGTGCATTGTCTGCATATCGAT 1714


RESULT   9
US-08-761-258-8/c
; Sequence 8, Application US/08761258
; Patent No. 5756087
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Lam, Stephen T.
;    APPLICANT:  Gaffney, Thomas D.
;    APPLICANT:  Torkewitz, Nancy
;    TITLE OF INVENTION:  Genetically Modified Pseudomonas Strains
;    TITLE OF INVENTION:  with Enhanced Biocontrol Activity
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/761,258
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3186 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas fluorescens
;      STRAIN:  CGA267356 (aka MOCG134 and aka BL915)
;    IMMEDIATE SOURCE:
;      CLONE:  pCIB146
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  245..251
;      OTHER INFORMATION:  /note= "potential ribosome binding
;      OTHER INFORMATION:  site"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  256..3006
;      OTHER INFORMATION:  /product= "LemA"
;      OTHER INFORMATION:  /note= "LemA coding sequence."
US-08-761-258-8


  Query Match             68.8%;  Score 17.2;  DB 1;  Length 3186;
  Best Local Similarity   86.4%;  Pred. No. 32;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 caggtgcatcgccagcatttcg 23
          |||| |||| |||||||| |||
Db   2763 CAGGAGCATGGCCAGCATGTCG 2742


RESULT  10
US-08-977-306-8/c
; Sequence 8, Application US/08977306
; Patent No. 5955348
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Gaffney, Thomas D.
;    APPLICANT:  Torkewitz, Nancy
;    APPLICANT:  Stafford, Jill M.
;    TITLE OF INVENTION:  Genetically Modified Pseudomonas Strains
;    TITLE OF INVENTION:  with Enhanced Biocontrol Activity
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5955348artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/977,306
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3186 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudomonas fluorescens
;      STRAIN:  CGA267356 (aka MOCG134 and aka BL915)
;    IMMEDIATE SOURCE:
;      CLONE:  pCIB146
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  245..251
;      OTHER INFORMATION:  /note= "potential ribosome binding
;      OTHER INFORMATION:  site"
```

```
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  256..3006
;      OTHER INFORMATION:  /product= "LemA"
;      OTHER INFORMATION:  /note= "LemA coding sequence."
US-08-977-306-8


  Query Match             68.8%;  Score 17.2;  DB 2;  Length 3186;
  Best Local Similarity   86.4%;  Pred. No. 32;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcg 23
           |||| |||| ||||||||| |||
Db    2763 CAGGAGCATGGCCAGCATGTCG 2742


RESULT  11
US-08-483-376-1
; Sequence 1, Application US/08483376
; Patent No. 5955330
;  GENERAL INFORMATION:
;    APPLICANT:  Vasil, Vimla
;    APPLICANT:  Clancy, Maureen A.
;    APPLICANT:  Ferl, Robert J.
;    APPLICANT:  Vasil, Indra K.
;    APPLICANT:  Hannah, L. C.
;    TITLE OF INVENTION:  No. 5955330el Means for Enhancing Gene
;    TITLE OF INVENTION:  Expression
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,376
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/418,540
;      FILING DATE:  07-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/102,115
;      FILING DATE:  04-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/830,956
;      FILING DATE:  05-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/353,854
;      FILING DATE:  18-MAY-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  10-94B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6386 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Zea mays
;      STRAIN:  Black Sweet
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  131..182
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1211..1324
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1828..1948
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  2041..2187
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  2269..2460
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  2605..2728
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  2822..3038
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  3256..3351
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  3447..3620
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  3702..3818
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  3912..4078
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  4158..4381
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  4517..4835
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  4768..5212
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  5372..5510
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  5636..5917
US-08-483-376-1


  Query Match             68.0%;  Score 17;  DB 2;  Length 6386;
  Best Local Similarity   80.0%;  Pred. No. 45;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           | ||   ||||||| ||||||||||
Db    5691 GGAGAACCATCGCCTGCATTTCGAT 5715


RESULT  12
US-08-861-774E-53/c
; Sequence 53, Application US/08861774E
; Patent No. 6297007
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
```

```
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/08/861,774E
;  CURRENT FILING DATE:  1997-05-22
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
;   LENGTH: 643
;   TYPE: DNA
;   ORGANISM: Peltigera neopolydactyla
US-08-861-774E-53


  Query Match             67.2%;  Score 16.8;  DB 4;  Length 643;
  Best Local Similarity   90.0%;  Pred. No. 38;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcattt 21
           ||||||| | ||||||||||
Db      53 CAGGTGCTTTGCCAGCATTT 34


RESULT  13
US-08-833-167-19
; Sequence 19, Application US/08833167
; Patent No. 6100070
;  GENERAL INFORMATION:
;    APPLICANT:  ZURFLUH, LINDA L
;    APPLICANT:  MCWHERTER, CHARLES A
;    APPLICANT:  MCKEARN, JOHN P
;    APPLICANT:  KLEIN, BARBARA K
;    APPLICANT:  FENG, YIQING
;    APPLICANT:  BRAFORD-GOLDBERG, SARAH R
;    APPLICANT:  LEE, STEPHEN C
;    TITLE OF INVENTION:  G-CSF RECEPTOR AGONISTS
;    NUMBER OF SEQUENCES:  129
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DENNIS A BENNETT, G.D. SEARLE & CO.,
;      ADDRESSEE:  CORPORATE PATENT DEPT.
;      STREET:  P.O. BOX 5110
;      CITY:  CHICAGO
;      STATE:  ILLINOIS
;      COUNTRY:  USA
;      ZIP:  60680
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/833,167
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US 96/15935
;      FILING DATE:  04-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/004,382
;      FILING DATE:  05-OCT-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENNETT, DENNIS A
;      REFERENCE/DOCKET NUMBER:  2907/1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  314-694-5402
;      TELEFAX:  314-694-9095
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  65 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA (synthetic)"
US-08-833-167-19


  Query Match             66.4%;  Score 16.6;  DB 3;  Length 65;
  Best Local Similarity   82.6%;  Pred. No. 33;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 aggtgcatcgccagcatttcgat 25
           |||||||  ||||||||||  ||
Db      40 AGGTGCAATGCCAGCATTTGCAT 62


RESULT  14
US-08-833-167-40
; Sequence 40, Application US/08833167
; Patent No. 6100070
;  GENERAL INFORMATION:
;    APPLICANT:  ZURFLUH, LINDA L
;    APPLICANT:  MCWHERTER, CHARLES A
;    APPLICANT:  MCKEARN, JOHN P
;    APPLICANT:  KLEIN, BARBARA K
;    APPLICANT:  FENG, YIQING
;    APPLICANT:  BRAFORD-GOLDBERG, SARAH R
;    APPLICANT:  LEE, STEPHEN C
;    TITLE OF INVENTION:  G-CSF RECEPTOR AGONISTS
;    NUMBER OF SEQUENCES:  129
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DENNIS A BENNETT, G.D. SEARLE & CO.,
;      ADDRESSEE:  CORPORATE PATENT DEPT.
;      STREET:  P.O. BOX 5110
;      CITY:  CHICAGO
;      STATE:  ILLINOIS
;      COUNTRY:  USA
;      ZIP:  60680
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/833,167
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US 96/15935
;      FILING DATE:  04-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/004,382
;      FILING DATE:  05-OCT-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENNETT, DENNIS A
;      REFERENCE/DOCKET NUMBER:  2907/1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  314-694-5402
;      TELEFAX:  314-694-9095
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  522 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA (synthetic)"
US-08-833-167-40


  Query Match             66.4%;  Score 16.6;  DB 3;  Length 522;
  Best Local Similarity   82.6%;  Pred. No. 46;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 aggtgcatcgccagcatttcgat 25
           |||||||  ||||||||||  ||
```

```
Db      33 AGGTGCAATGCCAGCATTTGCAT 55


RESULT  15
US-08-833-167-36
; Sequence 36, Application US/08833167
; Patent No. 6100070
;  GENERAL INFORMATION:
;     APPLICANT:  ZURFLUH, LINDA L
;     APPLICANT:  MCWHERTER, CHARLES A
;     APPLICANT:  MCKEARN, JOHN P
;     APPLICANT:  KLEIN, BARBARA K
;     APPLICANT:  FENG, YIQING
;     APPLICANT:  BRAFORD-GOLDBERG, SARAH R
;     APPLICANT:  LEE, STEPHEN C
;     TITLE OF INVENTION:  G-CSF RECEPTOR AGONISTS
;     NUMBER OF SEQUENCES:  129
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  DENNIS A BENNETT, G.D. SEARLE & CO.,
;       ADDRESSEE:  CORPORATE PATENT DEPT.
;       STREET:  P.O. BOX 5110
;       CITY:  CHICAGO
;       STATE:  ILLINOIS
;       COUNTRY:  USA
;       ZIP:  60680
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/833,167
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  WO PCT/US 96/15935
;       FILING DATE:  04-OCT-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/004,382
;       FILING DATE:  05-OCT-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENNETT, DENNIS A
;       REFERENCE/DOCKET NUMBER:  2907/1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  314-694-5402
;       TELEFAX:  314-694-9095
;   INFORMATION FOR SEQ ID NO:  36:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  531 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
;       DESCRIPTION:  /desc = "DNA (synthetic)"
US-08-833-167-36


  Query Match             66.4%;  Score 16.6;  DB 3;  Length 531;
  Best Local Similarity   82.6%;  Pred. No. 46;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 aggtgcatcgccagcatttcgat 25
           |||||||  ||||||||||  ||
Db      33 AGGTGCAATGCCAGCATTTGCAT 55


RESULT  16
US-07-642-734C-3/c
; Sequence 3, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;     APPLICANT:  Katz, L
;     APPLICANT:  Donadio, S
;     APPLICANT:  Mcalpine, J B
;     TITLE OF INVENTION:  Recombinant DNA Method for Producing
;     TITLE OF INVENTION:  Erythromycin Analogs
;     NUMBER OF SEQUENCES:  27
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Edward H. Gorman
;       STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;       STREET:  Park Rd
;       CITY:  Abbott Park
;       STATE:  IL
;       COUNTRY:  US
;       ZIP:  60064-3500
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/642,734C
;       FILING DATE:  17-JAN-91
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Danckers, Andreas M
;       REGISTRATION NUMBER:  32652
;       REFERENCE/DOCKET NUMBER:  4952.US.01
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  708-937-9396
;       TELEFAX:  708-938-2623
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  20235 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Saccharopolyspora erythraea
;       STRAIN:  NRRL 238
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  19..10722
;       OTHER INFORMATION:  /codon_start= 19
;       OTHER INFORMATION:  /function= "gene eryA"
;       OTHER INFORMATION:  /product= "eryA ORF2 encoding modules 3 & 4 for
;       OTHER INFORMATION:  6-deoxyerythronolide B"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19..4470
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  97..1482
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacyl ACP synthase of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1693..2670
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3406..3921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4171..4428
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 3"
;     FEATURE:
```

```
;       NAME/KEY:  misc_feature
;       LOCATION:  4471..10722
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4471..5847
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynhase domain of module"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6054..7026
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  7165..9216
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  dehydratase and enoylreductase domains m"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9433..9984
;       OTHER INFORMATION:  /function= "approximate span
;       OTHER INFORMATION:  beta-ketoreductase of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10225..10483
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 4"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  10723..20235
;       OTHER INFORMATION:  /codon_start= 10723
;       OTHER INFORMATION:  /function= "gene =eryA"
;       OTHER INFORMATION:  /product= "orf3 encoding modules 5 & 6
;       OTHER INFORMATION:  6-deoxyerythronolide B formatio"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10723..15165
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10831..12174
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  12379..13350
;       OTHER INFORMATION:  /function= "approximatr span of
;       OTHER INFORMATION:  acyltransferase domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14062..14610
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14857..15114
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15166..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15172..16569
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  16768..17721
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  18379..18921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19149..19398
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19492..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  thioesterase domain of module 6"
US-07-642-734C-3


  Query Match             66.4%;  Score 16.6;  DB 1;  Length 20235;
  Best Local Similarity   82.6%;  Pred. No. 82;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcg 23
           || ||| |||||||||| ||||
Db    4750 GCGGGTCCATCGCCAGCGCTTCG 4728


RESULT  17
US-08-439-009A-3/c
; Sequence 3, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20235 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
```

```
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 238
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  19..10722
;      OTHER INFORMATION:  /codon_start= 19
;      OTHER INFORMATION:  /function= "gene eryA"
;      OTHER INFORMATION:  /product= "eryA ORF2 encoding modules 3 & 4 for
;      OTHER INFORMATION:  6-deoxyerythronolide B"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19..4470
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  97..1482
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACP synthase of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1693..2670
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3406..3921
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4171..4428
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4471..10722
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4471..5847
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACPsynhase domain of module"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6054..7026
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7165..9216
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  dehydratase and enoylreductase domains m"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9433..9984
;      OTHER INFORMATION:  /function= "approximate span
;      OTHER INFORMATION:  beta-ketoreductase of module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10225..10483
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module !"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  10723..20235
;      OTHER INFORMATION:  /codon_start= 10723
;      OTHER INFORMATION:  /function= "gene =eryA"
;      OTHER INFORMATION:  /product= "orf3 encoding modules 5 & 6
;      OTHER INFORMATION:  6-deoxyerythronolide B formatio"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10723..15165
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 5"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10831..12174
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  12379..13350
;      OTHER INFORMATION:  /function= "approximatr span of
;      OTHER INFORMATION:  acyltransferase domain of module 5"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  14062..14610
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase of module 5"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  14857..15114
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 5"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  15166..20235
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 6"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  15172..16569
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  16768..17721
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 6"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  18379..18921
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 6"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19149..19398
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 6"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19492..20235
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  thioesterase domain of module 6"
US-08-439-009A-3


  Query Match             66.4%;  Score 16.6;  DB 3;  Length 20235;
  Best Local Similarity   82.6%;  Pred. No. 82;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcg 23
           || ||| ||||||||||  ||||
Db    4750 GCGGGTCCATCGCCAGCGCTTCG 4728


RESULT  18
US-08-759-444-2
; Sequence 2, Application US/08759444
; Patent No. 5824309
;  GENERAL INFORMATION:
;    APPLICANT:  DasSarma, Shiladitya
;    APPLICANT:  Morshed, Fazeela
```

```
;    APPLICANT:  Stuart, Elizabeth
;    APPLICANT:  Black, Samuel
;    TITLE OF INVENTION:  RECOMBINANT GAS VESICLES AND USES THEREOF
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/759,444
;      FILING DATE:  05-DEC-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/008,200
;      FILING DATE:  05-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meiklejohn, Ph.D., Anita L
;      REGISTRATION NUMBER:  35,283
;      REFERENCE/DOCKET NUMBER:  07880/003001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-542-5070
;      TELEFAX:  617-542-8906
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8878 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-759-444-2


  Query Match             64.8%;  Score 16.2;  DB 1;  Length 8878;
  Best Local Similarity   85.7%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 aggtgcatcgccagcatttcg 23
           ||| ||||| ||||||| |||
Db    1602 AGGAGCATCTCCAGCATCTCG 1622


RESULT  19
US-08-680-897-1
; Sequence 1, Application US/08680897
; Patent No. 6008051
;  GENERAL INFORMATION:
;    APPLICANT:  DasSarma, Shiladitya
;    APPLICANT:  Halladay, John
;    APPLICANT:  Ng, Wai-lap
;    TITLE OF INVENTION:  RECOMBINANT VECTOR AND PROCESS FOR CELL
;    TITLE OF INVENTION:  FLOTATION
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Scully, Scott, Murphy & Presser
;      STREET:  400 Garden City Plaza
;      CITY:  Garden City
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  11530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/680,897
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/271,270
;      FILING DATE:
;      APPLICATION NUMBER:  944,581
;      FILING DATE:  14-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DiGiglio, Frank S.
;      REGISTRATION NUMBER:  31,346
;      REFERENCE/DOCKET NUMBER:  8680
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (516) 742-4343
;      TELEFAX:  (516) 742-4366
;      TELEX:  230 901 SANS UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9880 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-680-897-1


  Query Match             64.8%;  Score 16.2;  DB 3;  Length 9880;
  Best Local Similarity   85.7%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 aggtgcatcgccagcatttcg 23
           ||| ||||| ||||||| |||
Db    2602 AGGAGCATCTCCAGCATCTCG 2622


RESULT  20
US-09-167-717-3/c
; Sequence 3, Application US/09167717
; Patent No. 6133014
; GENERAL INFORMATION:
;  APPLICANT: MUKOUYAMA, MASAHARU
;  APPLICANT: YASUDA, SHINZO
;  APPLICANT: KOMATSUZAKI, SATOMI
;  TITLE OF INVENTION: MALEATE ISOMERASE GENE
;  FILE REFERENCE: PH-555US
;  CURRENT APPLICATION NUMBER: US/09/167,717
;  CURRENT FILING DATE: 1998-10-07
;  EARLIER APPLICATION NUMBER: JP 276261/1997
;  EARLIER FILING DATE: 1997-10-08
;  NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 3
;   LENGTH: 750
;   TYPE: DNA
;   ORGANISM: Pseudomonas fluorescens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(750)
US-09-167-717-3


  Query Match             64.0%;  Score 16;  DB 3;  Length 750;
  Best Local Similarity   79.2%;  Pred. No. 92;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcgat 25
           |||| |||||||| | || ||| |
Db      81 CAGGAGCATCGCCGGGATCTCGGT 58


RESULT  21
US-08-687-865A-1/c
; Sequence 1, Application US/08687865A
; Patent No. 5955596
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Jones, Kevin F.
;    APPLICANT:  Zagursky, Robert J.
;    APPLICANT:  Ooi, Peggy
;    TITLE OF INVENTION:  The NucA Protein of Haemophilus
;    TITLE OF INVENTION:  influenzae and the Gene Encoding That Protein
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  American Cyanamid Company
;      STREET:  One Cyanamid Plaza
;      CITY:  Wayne
;      STATE:  New Jersey
;      COUNTRY:  U.S.A.
;      ZIP:  07470
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/687,865A
;      FILING DATE:  26-JUL-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gordon, Alan M.
;      REGISTRATION NUMBER:  30,637
;      REFERENCE/DOCKET NUMBER:  33,250-00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-831-3244
;      TELEFAX:  201-831-3305
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2263 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  229..2037
US-08-687-865A-1


  Query Match             64.0%;  Score 16;  DB 2;  Length 2263;
  Best Local Similarity   79.2%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcga 24
           ||  || || ||||||||||| ||
Db    1541 GCGCGTACACCGCCAGCATTTTGA 1518


RESULT  22
US-09-043-711-1/c
; Sequence 1, Application US/09043711
; Patent No. 6221365
;  GENERAL INFORMATION:
;    APPLICANT:  Jones, Kevin F.
;    APPLICANT:  Zagursky, Robert J.
;    APPLICANT:  Ooi, Peggy
;    TITLE OF INVENTION:  The NucA Protein of Haemophilus
;    TITLE OF INVENTION:  influenzae and the Gene Encoding That Protein
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  American Cyanamid Company
;      STREET:  One Cyanamid Plaza
;      CITY:  Wayne
;      STATE:  New Jersey
;      COUNTRY:  U.S.A.
;      ZIP:  07470
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/043,711
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/687,865
;      FILING DATE:  26-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gordon, Alan M.
;      REGISTRATION NUMBER:  30,637
;      REFERENCE/DOCKET NUMBER:  33,250-00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-831-3244
;      TELEFAX:  201-831-3305
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2263 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  229..2037
US-09-043-711-1


  Query Match             64.0%;  Score 16;  DB 4;  Length 2263;
  Best Local Similarity   79.2%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcga 24
           ||  || || ||||||||||| ||
Db    1541 GCGCGTACACCGCCAGCATTTTGA 1518


RESULT  23
US-08-866-757-1
; Sequence 1, Application US/08866757
; Patent No. 5858716
;  GENERAL INFORMATION:
;    APPLICANT:  ELSHOURBAGY, NABIL A
;    APPLICANT:  LI, XIAOTONG
;    APPLICANT:  BERGSMA, DERK J
;    TITLE OF INVENTION:  NOVEL 7TM RECEPTOR (H2CAA71)
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/866,757
;      FILING DATE:  30-MAY-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH-70055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
```

```
;       TELEX:  846169
;    INFORMATION FOR SEQ ID NO:  1:
;      SEQUENCE CHARACTERISTICS:
;        LENGTH:  4203 base pairs
;        TYPE:  nucleic acid
;        STRANDEDNESS:  single
;        TOPOLOGY:  linear
;      MOLECULE TYPE:  cDNA
US-08-866-757-1


  Query Match             64.0%;  Score 16;  DB 2;  Length 4203;
  Best Local Similarity   79.2%;  Pred. No. 1.2e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcgat 25
           |||||| | ||||| |||| |||
Db    1655 CAGGTGAAACGCCATCATTAGGAT 1678


RESULT  24
US-09-153-593-1
; Sequence 1, Application US/09153593A
; Patent No. 6174994
; GENERAL INFORMATION:
;  APPLICANT: ELSHOURBAGY, NABIL A
;  APPLICANT: LI, XIAOTONG
;  APPLICANT: BERGSMA, DERK J
;  TITLE OF INVENTION: NOVEL 7TM RECEPTOR (H2CAA71)
;  FILE REFERENCE: GH-70055-1
;  CURRENT APPLICATION NUMBER: US/09/153,593A
;  CURRENT FILING DATE: 1998-09-15
;  EARLIER APPLICATION NUMBER: 08/866,757
;  EARLIER FILING DATE: 1997-05-30
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 4203
;   TYPE: DNA
;   ORGANISM: HOMO SAPIENS
US-09-153-593-1


  Query Match             64.0%;  Score 16;  DB 4;  Length 4203;
  Best Local Similarity   79.2%;  Pred. No. 1.2e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcgat 25
           |||||| | ||||| |||| |||
Db    1655 caggtgaaacgccatcattaggat 1678


RESULT  25
US-07-853-913-1
; Sequence 1, Application US/07853913
; Patent No. 5338839
;  GENERAL INFORMATION:
;    APPLICANT:  McKay, Ronald D.G.
;    APPLICANT:  Lendahl, Urban
;    TITLE OF INVENTION:  Nestin Expression As An Indicator of
;    TITLE OF INVENTION:  Neuroepithelial Tumors
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/853,913
;      FILING DATE:  19920319
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/660,412
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/603,803
;      FILING DATE:  25-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/201,762
;      FILING DATE:  02-JUN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/180,548
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-4641AAAA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11236 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-07-853-913-1


  Query Match             64.0%;  Score 16;  DB 1;  Length 11236;
  Best Local Similarity   79.2%;  Pred. No. 1.4e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcga 24
           ||||| ||||| | |  |||||||
Db    1819 GCAGGAGCATCTCTACGATTTCGA 1842


RESULT  26
US-08-125-468-1
; Sequence 1, Application US/08125468
; Patent No. 5589385
;  GENERAL INFORMATION:
;    APPLICANT:  Ryan, Michael J.
;    APPLICANT:  Lotvin, Jason A.
;    APPLICANT:  Strathy, Nancy
;    APPLICANT:  Fantini, Susan E.
;    TITLE OF INVENTION:  Cloning of the biosynthetic pathway for
;    TITLE OF INVENTION:  chlortetracycline and tetracyline Formation and cosmids
;    TITLE OF INVENTION:  useful therein
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  American Cyanamid Company
;      STREET:  One Cyanamid Plaza
;      CITY:  Wayne
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07470
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/125,468
;      FILING DATE:  22-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsevdos, Estelle J
```

```
;      REGISTRATION NUMBER:  31,145
;      REFERENCE/DOCKET NUMBER:  31,255-02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201)831-3241
;      TELEFAX:  (201)831-3305
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30001 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-125-468-1


  Query Match             64.0%;  Score 16;  DB 1;  Length 30001;
  Best Local Similarity   79.2%;  Pred. No. 1.6e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcga 24
          ||||| |||  |||||||  ||||
Db  28819 GCAGGAGCACGGCCAGCAGATCGA 28842


RESULT  27
US-08-474-933-1
; Sequence 1, Application US/08474933
; Patent No. 5866410
;  GENERAL INFORMATION:
;    APPLICANT:  Ryan, Michael J.
;    APPLICANT:  Lotvin, Jason A.
;    APPLICANT:  Strathy, Nancy
;    APPLICANT:  Fantini, Susan E.
;    TITLE OF INVENTION:  Cloning of the biosynthetic pathway for
;    TITLE OF INVENTION:  chlortetracycline and tetracyline Formation and cosmids
;    TITLE OF INVENTION:  useful therein
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  American Cyanamid Company
;      STREET:  One Cyanamid Plaza
;      CITY:  Wayne
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07470
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,933
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/125,468
;      FILING DATE:  22-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsevdos, Estelle J
;      REGISTRATION NUMBER:  31,145
;      REFERENCE/DOCKET NUMBER:  31,255-02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201)831-3241
;      TELEFAX:  (201)831-3305
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  30001 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-474-933-1


  Query Match             64.0%;  Score 16;  DB 2;  Length 30001;
  Best Local Similarity   79.2%;  Pred. No. 1.6e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcga 24
          ||||| |||  |||||||  ||||
Db  28819 GCAGGAGCACGGCCAGCAGATCGA 28842


RESULT  28
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


  Query Match             64.0%;  Score 16;  DB 4;  Length 4403765;
  Best Local Similarity   86.4%;  Pred. No. 50;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttc 22
          |||||||||||||||| || | |
Db 2873375 gcaggtgcatcgccaccagtgc 2873396


RESULT  29
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1


  Query Match             64.0%;  Score 16;  DB 4;  Length 4411529;
  Best Local Similarity   86.4%;  Pred. No. 50;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy       1 gcaggtgcatcgccagcatttc 22
           ||||||||||||||| || | |
Db 2877266 gcaggtgcatcgccaccagtgc 2877287


RESULT  30
US-08-620-467A-9/c
; Sequence 9, Application US/08620467A
; Patent No. 5798231
;  GENERAL INFORMATION:
;    APPLICANT:  HERCEND, THIERRY; TRIEBEL,
;    APPLICANT:  FREDERIC; ROMAN-ROMAN, SERGIO; FERRADINI,
;    APPLICANT:  LAURENT
;    TITLE OF INVENTION:  NUCLEOTIDE SEQUENCE CODING
;    TITLE OF INVENTION:  FOR VARIABLE REGIONS OF THE ' CHAINS OF HUMAN T
;    TITLE OF INVENTION:  LYMPHOCYTE RECEPTORS, CORRESPONDING PEPTIDE
;    TITLE OF INVENTION:  SEGMENTS AND DIAGNOSTIC AND THERAPEUTIC USES
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIERMAN, MUSERLIAN AND LUCAS,
;      ADDRESSEE:  LLP
;      STREET:  600 THIRD AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/620,467A
;      FILING DATE:  22-MAR-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/348,572
;      FILING DATE:  19-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/934,529
;      FILING DATE:  24-NOV-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR92/00111
;      FILING DATE:  07-FEB-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CHARLES A. MUSERLIAN
;      REGISTRATION NUMBER:  19,683
;      REFERENCE/DOCKET NUMBER:  146.1159-CON-DIV-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-661-8000
;      TELEFAX:  212-661-8002
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  386
;      TYPE:  NUCLEOTIDE
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  HUMAN
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      HAPLOTYPE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:  HUMAN T LYMPHOCYTE
;      ORGANELLE:
;    FEATURE:
;      NAME/KEY:  IGR a 11
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:  SEQUENCE V   7

US-08-620-467A-9


  Query Match             62.4%;  Score 15.6;  DB 1;  Length 386;
  Best Local Similarity   81.8%;  Pred. No. 1.3e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       4 ggtgcatcgccagcatttcgat 25
           ||||| |||||||||| | | |
Db     247 GGTGCTTCGCCAGCATGTTGCT 226


RESULT  31
US-08-348-572-9/c
; Sequence 9, Application US/08348572
; Patent No. 5817511
;  GENERAL INFORMATION:
;    APPLICANT:  HERCEND, THIERRY; TRIEBEL, FREDERIC;
;    APPLICANT:  ROMAN-ROMAN, SERGIO; FERRADINI, LAURENT
;    TITLE OF INVENTION:  NUCLEOTIDE SEQUENCE CODING FOR VARIABLE
;    TITLE OF INVENTION:  REGIONS OF THE ALPHA CHAINS OF HUMAN T LYMPHOCYTE RECEPTORS
;    TITLE OF INVENTION:  CORRESPONDING PEPTIDE SEGMENTS AND THE DIAGNOSTIC AND
;    TITLE OF INVENTION:  THERAPEUTIC USES
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIERMAN & MUSERLIAN
;      STREET:  600 THIRD AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/348,572
;      FILING DATE:  19-APR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/934,529
;      FILING DATE:  24-NOV-1992
;      APPLICATION NUMBER:  PCT/FR 92/00111
;      FILING DATE:  07-FEB-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CHARLES A. MUSERLIAN
;      REGISTRATION NUMBER:  19,683
;      REFERENCE/DOCKET NUMBER:  146.1159
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-661-8000
;      TELEFAX:  212-661-8002
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  386
;      TYPE:  NUCLEOTIDE
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  HUMAN
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      HAPLOTYPE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:  HUMAN T LYMPHOCYTE
;      ORGANELLE:
;    FEATURE:
;      NAME/KEY:  IGR a 11
;      LOCATION:
;      IDENTIFICATION METHOD:
```

```
;      OTHER INFORMATION:  SEQUENCE V Alpha 7
US-08-348-572-9


  Query Match             62.4%;  Score 15.6;  DB 1;  Length 386;
  Best Local Similarity   81.8%;  Pred. No. 1.3e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       4 ggtgcatcgccagcatttcgat 25
           ||||| |||||||||| | | |
Db     247 GGTGCTTCGCCAGCATGTTGCT 226


RESULT  32
US-09-041-090B-9/c
; Sequence 9, Application US/09041090B
; Patent No. 6114516
;  GENERAL INFORMATION:
;    APPLICANT:  HERCEND, THIERRY; TRIEBEL, FREDERIC;
;    APPLICANT:  ROMAN-ROMAN, SERGIO; FERRADINI, LAURENT
;    TITLE OF INVENTION:  NUCLEOTIDE SEQUENCE CODING FOR VARIABLE
;    TITLE OF INVENTION:  REGIONS OF THE ALPHA CHAINS OF HUMAN T LYMPHOCYTE RECEPTORS,
;    TITLE OF INVENTION:  CORRESPONDING PEPTIDE SEGMENTS AND THE DIAGNOSTIC AND
;    TITLE OF INVENTION:  THERAPEUTIC USES
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIERMAN & MUSERLIAN
;      STREET:  600 THIRD AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,090B
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,572
;      FILING DATE:  19-APR-1995
;      APPLICATION NUMBER:  PCT/FR 92/00111
;      FILING DATE:  07-FEB-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CHARLES A. MUSERLIAN
;      REGISTRATION NUMBER:  19,683
;      REFERENCE/DOCKET NUMBER:  146.1159
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-661-8000
;      TELEFAX:  212-661-8002
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  386
;      TYPE:  NUCLEOTIDE
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  HUMAN
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      HAPLOTYPE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:  HUMAN T LYMPHOCYTE
;      ORGANELLE:
;    FEATURE:
;      NAME/KEY:  IGR a 11
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:  SEQUENCE V Alpha 7
US-09-041-090B-9


  Query Match             62.4%;  Score 15.6;  DB 3;  Length 386;
  Best Local Similarity   81.8%;  Pred. No. 1.3e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       4 ggtgcatcgccagcatttcgat 25
           ||||| |||||||||| | | |
Db     247 GGTGCTTCGCCAGCATGTTGCT 226


RESULT  33
US-08-998-416-315/c
; Sequence 315, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  315:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  827 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1255RP
US-08-998-416-315


  Query Match             62.4%;  Score 15.6;  DB 4;  Length 827;
  Best Local Similarity   81.8%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcg 23
           || |  |||||||||||| ||||
```

```
Db    250 CACGCCCATCGCCAGCAGTTCG 229


RESULT  34
US-08-233-788A-48/c
; Sequence 48, Application US/08233788A
; Patent No. 5635617
;  GENERAL INFORMATION:
;    APPLICANT:  Doran, James L.
;    APPLICANT:  Kay, William W.
;    APPLICANT:  Collinson, Karen S.
;    APPLICANT:  Clouthier, Sharon C.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR DETECTION
;    TITLE OF INVENTION:  OF SALMONELLA
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,788A
;      FILING DATE:  26-APR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  King, Joshua
;      REGISTRATION NUMBER:  35,570
;      REFERENCE/DOCKET NUMBER:  920043.403C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;      TELEX:  3723836 SEEDANBERRY
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1126 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(3..23, 27..944, 948..1124)
US-08-233-788A-48


  Query Match             62.4%;  Score 15.6;  DB 1;  Length 1126;
  Best Local Similarity   81.8%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttc 22
           |||||| |  | ||||||||||
Db     789 GCAGGTTCTGCTCCAGCATTTC 768


RESULT  35
5196511-1/c
;Patent No. 5196511
;    APPLICANT: PLOW, EDWARD F.;D'SOUZA, STANLEY E.
;GINSBERG,MARK H.
;    TITLE OF INVENTION: PEPTIDES AND ANTIBODIES THAT INHIBIT
;INTEGRIN-LIGAND BINDING
;    NUMBER OF SEQUENCES: 31
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/444,777
;      FILING DATE: 01-DEC-1989
;SEQ ID NO:1:
;      LENGTH: 3172
5196511-1


  Query Match             62.4%;  Score 15.6;  DB 6;  Length 3172;
  Best Local Similarity   81.8%;  Pred. No. 1.8e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttc 22
           |||||||||| |||||||  | |
Db    2054 GCAGGTGCACCGCCAGCTCTGC 2033


RESULT  36
US-09-060-756-1
; Sequence 1, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
;  APPLICANT: Cole, Stewart
;  APPLICANT: Buchrieser-Brosch, Roland
;  APPLICANT: Gordon, Stephen
;  APPLICANT: Billault, Alain
;  TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
;  TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
;  TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
;  FILE REFERENCE: 3495-0169
;  CURRENT APPLICATION NUMBER: US/09/060,756
;  CURRENT FILING DATE: 1998-04-16
;  NUMBER OF SEQ ID NOS: 743
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 12732
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-09-060-756-1


  Query Match             62.4%;  Score 15.6;  DB 4;  Length 12732;
  Best Local Similarity   81.8%;  Pred. No. 2.2e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       4 ggtgcatcgccagcatttcgat 25
           | |||| |||||||||  ||||
Db  11638 gctgcagcgccagcatggcgat 11659


RESULT  37
5208144-34
;Patent No. 5208144
;    APPLICANT: SMITH, JOHN A.;RAYCHOWDHURY, RAKTIMA;NILES,JOHN L.
;    TITLE OF INVENTION: METHOD FOR DETECTION OF HUMAN DNA
;CONTAINING THE GENE ENCODING LOW DENSITY LIPOPROTEIN RECEPTOR
;    NUMBER OF SEQUENCES: 42
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/396,697
;      FILING DATE: 22-AUG-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 313,682
;      FILING DATE: 22-FEB-1989
;      APPLICATION NUMBER: 235,211
;      FILING DATE: 23-AUG-1988
;SEQ ID NO:34:
;      LENGTH: 480
5208144-34


  Query Match             61.6%;  Score 15.4;  DB 6;  Length 480;
  Best Local Similarity   76.0%;  Pred. No. 1.6e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||||||| |  ||| || | |
Db     295 gcaggtgcatctcaggcaattgggt 319
```

```
RESULT  38
US-08-998-416-155/c
; Sequence 155, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  155:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  644 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1059RP
US-08-998-416-155


  Query Match             61.6%;  Score 15.4;  DB 4;  Length 644;
  Best Local Similarity   76.0%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           ||| |||||||  || | |||| ||
Db     541 GCAAGTGCATCCACACCTTTTCCAT 517


RESULT  39
US-09-385-028-16
; Sequence 16, Application US/09385028
; Patent No. 6232106
;  GENERAL INFORMATION:
;    APPLICANT:  Susan E. Jensen
;    APPLICANT:  Kwamena A Aidoo
;    APPLICANT:  Ashish S. Paradkar
;    TITLE OF INVENTION:  DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
;    TITLE OF INVENTION:  Acid Biosynthesis
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JACOBSON, PRICE, HOLMAN & STERN,PPLC
;      STREET:  The Jenifer Buliding, 400 Seventh Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/385,028
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/790,462
;      FILING DATE:  29-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  D. Douglas Price
;      REGISTRATION NUMBER:  24,514
;      REFERENCE/DOCKET NUMBER:  1418/P57452US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202 638-6666
;      TELEFAX:  (202) 39305350
;      TELEX:  RCA 248593 IDEA UR
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1542 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-385-028-16


  Query Match             61.6%;  Score 15.4;  DB 4;  Length 1542;
  Best Local Similarity   76.0%;  Pred. No. 2e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||||||  ||| |||  ||||||
Db    1425 GCAGGTGGTGCGCGAGCTGTTCGAT 1449


RESULT  40
US-08-724-394A-18
; Sequence 18, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237el
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/724,394A
;       FILING DATE:  01-OCT-1996
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fitts, Renee A.
;       REGISTRATION NUMBER:  35,136
;       REFERENCE/DOCKET NUMBER:  017957-000100
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-576-0200
;       TELEFAX:  415-576-0300
;   INFORMATION FOR SEQ ID NO:  18:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2266 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  not relevant
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1..2266
;       OTHER INFORMATION:  /note= "cDNA 22B"
US-08-724-394A-18


  Query Match             61.6%;  Score 15.4;  DB 2;  Length 2266;
  Best Local Similarity   76.0%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           ||||| ||  | |||||||| |||
Db     971 GCAGGATCAGGGTCAGCATTTGGAT 995


RESULT  41
US-08-356-060A-42/c
; Sequence 42, Application US/08356060A
; Patent No. 5844079
;  GENERAL INFORMATION:
;    APPLICANT:  Ingham, Phillip W.
;    APPLICANT:  McMahon, Andrew P.
;    APPLICANT:  Tabin, Clifford J.
;    TITLE OF INVENTION:  Vertebrate Embryonic Pattern-Inducing
;    TITLE OF INVENTION:  Proteins and Uses Related Thereto
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII(text)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,060A
;      FILING DATE:  14-DEC-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/176,427
;      FILING DATE:  30-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vincent, Matthew P.
;      REGISTRATION NUMBER:  36,709
;      REFERENCE/DOCKET NUMBER:  HMI-006CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3900 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3897
US-08-356-060A-42


  Query Match             61.6%;  Score 15.4;  DB 2;  Length 3900;
  Best Local Similarity   76.0%;  Pred. No. 2.3e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||| ||| |||| || ||  ||||
Db    3492 GCAGCTGCCTCGCGAGGATCGCGAT 3468


RESULT  42
US-08-460-900C-42/c
; Sequence 42, Application US/08460900C
; Patent No. 6165747
;  GENERAL INFORMATION:
;    APPLICANT:  Ingham, Phillip W.
;    APPLICANT:  McMahon, Andrew P.
;    APPLICANT:  Tabin, Clifford J.
;    APPLICANT:  Bumcrot, David A.
;    APPLICANT:  Marti-Gorostiza, Elisa
;    TITLE OF INVENTION:  Vertebrate Embryonic Pattern-Inducing
;    TITLE OF INVENTION:  Proteins and Uses Related Thereto
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;      STREET:  One Post Office Square
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,900C
;      FILING DATE:  5-JUNE-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/435,093
;      FILING DATE:  4-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/356,060
;      FILING DATE:  14-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/176,427
;      FILING DATE:  30-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vincent, Matthew P.
;      REGISTRATION NUMBER:  36,709
;      REFERENCE/DOCKET NUMBER:  HMV-006.05
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 832-1000
;      TELEFAX:  (617) 832-7000
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3900 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3897
US-08-460-900C-42


  Query Match             61.6%;  Score 15.4;  DB 4;  Length 3900;
  Best Local Similarity   76.0%;  Pred. No. 2.3e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcgat 25
          |||| ||| |||| || ||  ||||
Db   3492 GCAGCTGCCTCGCGAGGATCGCGAT 3468


RESULT  43
US-08-674-509B-42/c
; Sequence 42, Application US/08674509B
; Patent No. 6261786
;  GENERAL INFORMATION:
;    APPLICANT:  Ingham, Phillip W.
;    APPLICANT:  McMahon, Andrew P.
;    APPLICANT:  Tabin, Clifford J.
;    APPLICANT:  Marigo, Valeria
;    TITLE OF INVENTION:  SCREENING ASSAYS FOR HEDGEGOG AGONISTS
;    TITLE OF INVENTION:  AND ANTAGONISTS
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;      STREET:  One Post Office Square
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109-2170
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/674,509B
;      FILING DATE:  02-JUL-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/460,900
;      FILING DATE:  05-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vincent, Matthew P.
;      REGISTRATION NUMBER:  36,709
;      REFERENCE/DOCKET NUMBER:  HMV-006.06
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-832-1000
;      TELEFAX:  617-832-7000
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3900 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3897
US-08-674-509B-42


  Query Match             61.6%;  Score 15.4;  DB 4;  Length 3900;
  Best Local Similarity   76.0%;  Pred. No. 2.3e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcgat 25
          |||| ||| |||| || ||  ||||
Db   3492 GCAGCTGCCTCGCGAGGATCGCGAT 3468


RESULT  44
US-08-954-698-42/c
; Sequence 42, Application US/08954698
; Patent No. 6271363
;  GENERAL INFORMATION:
;    APPLICANT:  Ingham, Phillip W.
;    APPLICANT:  McMahon, Andrew P.
;    APPLICANT:  Tabin, Clifford J.
;    TITLE OF INVENTION:  Vertebrate Embryonic Pattern-Inducing
;    TITLE OF INVENTION:  Proteins and Uses Related Thereto
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;      STREET:  One Post Office Square
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109-2170
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/954,698
;      FILING DATE:  20-OCT-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/462,386
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/435,093
;      FILING DATE:  04-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/356,060
;      FILING DATE:  14-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/176,427
;      FILING DATE:  30-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vincent, Matthew P.
;      REGISTRATION NUMBER:  36,709
;      REFERENCE/DOCKET NUMBER:  HMV-006.10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-832-1000
;      TELEFAX:  617-832-7000
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3900 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..3897
US-08-954-698-42


  Query Match             61.6%;  Score 15.4;  DB 4;  Length 3900;
  Best Local Similarity   76.0%;  Pred. No. 2.3e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcgat 25
          |||| ||| |||| || ||  ||||
Db   3492 GCAGCTGCCTCGCGAGGATCGCGAT 3468


RESULT  45
US-08-881-706-1/c
; Sequence 1, Application US/08881706
```

```
; Patent No. 6245969
; GENERAL INFORMATION:
;  APPLICANT: Chong, Joane
;  APPLICANT: Li, Jianming
;  TITLE OF INVENTION: Receptor Kinase BIN1
;  FILE REFERENCE: 07251/022001
;  CURRENT APPLICATION NUMBER: US/08/881,706
;  CURRENT FILING DATE: 1997-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 4104
;   TYPE: DNA
;   ORGANISM: Arabidopsis sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (97)..(3687)
US-08-881-706-1


  Query Match             61.6%;  Score 15.4;  DB 4;  Length 4104;
  Best Local Similarity   76.0%;  Pred. No. 2.3e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||| |  |||||||  |||| ||
Db    2624 GCAGTTCTATCGCCAGAGTTTCCAT 2600


Search completed: December 26, 2001, 12:25:34
Job time: 7291 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:11:42 ; Search time 4612.2 Seconds
(without alignments)
58.246 Million cell updates/sec

Title: US-09-396-196F-6
Perfect score: 25
Sequence: 1 gcaggtgcatcgccagcatttcgat 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	19.2	76.8	538	10	AV439422	AV439422 AV439422
	2	19.2	76.8	548	10	AV748431	AV748431 AV748431
	3	19.2	76.8	715	13	B07718	B07718 73611E03703
	4	19.2	76.8	801	13	AQ288588	AQ288588 nbxb0033K
c	5	18.6	74.4	376	10	AA879064	AA879064 oj18a12.s
c	6	18.6	74.4	627	11	BG455205	BG455205 NF102G09P
c	7	18.2	72.8	231	10	AI329089	AI329089 a9g12ne.r
	8	18.2	72.8	403	13	AQ849360	AQ849360 LMAJFV1_1
	9	18.2	72.8	527	13	P977L	AL390686 Leishmani
c	10	18.2	72.8	594	11	BI235362	BI235362 RE31201.5
c	11	18.2	72.8	597	11	BI228734	BI228734 RE26369.5
c	12	18.2	72.8	630	11	BI372325	BI372325 RE59263.5
c	13	18.2	72.8	659	11	BI228776	BI228776 RE26419.5
c	14	18.2	72.8	834	13	AZ176195	AZ176195 SP_0141_A
	15	17.8	71.2	360	13	P858R	AL390537 Leishmani
c	16	17.8	71.2	384	10	AV186249	AV186249 AV186249
c	17	17.8	71.2	439	13	AQ686803	AQ686803 nbxb0073I
c	18	17.8	71.2	498	10	AI532390	AI532390 SD03891.5
c	19	17.8	71.2	551	11	BI375275	BI375275 RE63047.5
c	20	17.8	71.2	580	13	AQ291570	AQ291570 nbxb0039K
	21	17.8	71.2	693	11	BF203823	BF203823 601868810
	22	17.8	71.2	733	11	BG718987	BG718987 602699291
c	23	17.8	71.2	887	11	BI160225	BI160225 602863918
	24	17.8	71.2	888	13	CNS003ZE	AL065360 Drosophil
c	25	17.8	71.2	1090	13	CNS058GS	AL325909 Tetraodon
c	26	17.8	71.2	1862	12	AK018497	AK018497 Mus muscu
	27	17.6	70.4	317	10	BE089771	BE089771 QV0-BT070
	28	17.6	70.4	411	11	BI371476	BI371476 RE58374.5
c	29	17.6	70.4	418	13	AQ890634	AQ890634 HS_2204_A
c	30	17.6	70.4	450	13	AQ516531	AQ516531 HS_5052_B
c	31	17.6	70.4	475	11	BF654469	BF654469 278665 MA
c	32	17.6	70.4	506	11	BG266805	BG266805 1000101G1
	33	17.6	70.4	518	11	BI423547	BI423547 949049D11
	34	17.6	70.4	549	10	AV619677	AV619677 AV619677
	35	17.6	70.4	558	13	AQ795581	AQ795581 nbxb0056A
c	36	17.6	70.4	582	10	AI665214	AI665214 605008C11
c	37	17.6	70.4	615	11	BF425225	BF425225 su54h10.y
c	38	17.6	70.4	944	13	CNS07883	AL433657 T7 end of
c	39	17.6	70.4	1032	13	CNS03VDY	AL262303 Tetraodon
	40	17.6	70.4	1039	13	CNS03B6D	AL236110 Tetraodon
c	41	17.6	70.4	1077	11	BG027526	BG027526 602295377
c	42	17.4	69.6	376	11	BF933544	BF933544 CM1-NT027
	43	17.4	69.6	527	13	AQ171736	AQ171736 HS_3172_B
	44	17.4	69.6	634	13	AQ350006	AQ350006 RPCI11-11
c	45	17.4	69.6	786	13	CNS03HMW	AL244481 Tetraodon
c	46	17.2	68.8	147	10	BE156980	BE156980 RC3-HT037
	47	17.2	68.8	309	10	AA686065	AA686065 EST109111
c	48	17.2	68.8	435	13	AZ273781	AZ273781 RPCI-23-1
	49	17.2	68.8	495	11	BF291841	BF291841 WHE2204_F
	50	17.2	68.8	519	13	TA47A09Q	AL455103 T. brucei
c	51	17.2	68.8	544	13	TA329B04Q	AL492386 T. brucei
	52	17.2	68.8	555	11	BF008069	BF008069 1698729 A
c	53	17.2	68.8	565	13	AZ212380	AZ212380 Sheared D
c	54	17.2	68.8	570	10	AA769145	AA769145 ob51h03.s
	55	17.2	68.8	639	11	BF503308	BF503308 AT19267.5
c	56	17.2	68.8	646	11	BF008062	BF008062 1664118 A
	57	17.2	68.8	687	11	BF493259	BF493259 AT01575.5
c	58	17.2	68.8	689	13	AQ641986	AQ641986 RPCI93-Ec
	59	17.2	68.8	699	11	BF493160	BF493160 AT01458.5
	60	17.2	68.8	708	10	BE513502	BE513502 601315005
	61	17.2	68.8	716	11	BF506001	BF506001 AT08347.5
	62	17.2	68.8	731	11	BF499903	BF499903 AT14681.5
	63	17.2	68.8	752	11	BF494889	BF494889 AT03660.5
	64	17.2	68.8	757	11	BF502685	BF502685 AT18484.5
	65	17.2	68.8	766	11	BF500581	BF500581 AT15567.5
	66	17.2	68.8	779	11	BE908400	BE908400 601503023
	67	17.2	68.8	780	11	BF347481	BF347481 602020507
	68	17.2	68.8	781	11	BG417267	BG417267 HVSMEk001
	69	17.2	68.8	810	11	BF493804	BF493804 AT02159.5
c	70	17.2	68.8	869	11	BG739775	BG739775 602630493
c	71	17.2	68.8	907	13	CNS01NTI	AL152663 Anopheles
c	72	17.2	68.8	977	11	BF624459	BF624459 HVSMEa001
	73	17.2	68.8	1039	13	CNS04XGW	AL311657 Tetraodon
	74	17.2	68.8	1101	13	CNS002LL	AL062860 Drosophil
c	75	17	68.0	135	11	C41614	C41614 C41614 Yuji
c	76	17	68.0	171	10	AI202275	AI202275 qs70b05.x
c	77	17	68.0	240	11	C49224	C49224 C49224 Yuji
c	78	17	68.0	281	10	AI978168	AI878168 614041A04
	79	17	68.0	299	10	AW355901	AW355901 707015F11
c	80	17	68.0	307	10	AI617514	AI617514 zehn1712.
c	81	17	68.0	328	10	AA173326	AA173326 zp02f03.r
c	82	17	68.0	360	10	AV185702	AV185702 AV185702
c	83	17	68.0	360	10	AV192272	AV192272 AV192272
c	84	17	68.0	360	10	AV196352	AV196352 AV196352
c	85	17	68.0	360	11	C41598	C41598 C41598 Yuji

c	86	17	68.0	360	11	C50371	C50371 C50371 Yuji
c	87	17	68.0	360	11	C68400	C68400 C68400 Yuji
c	88	17	68.0	360	11	C68621	C68621 C68621 Yuji
	89	17	68.0	362	10	AA274371	AA274371 TgESTzz25
c	90	17	68.0	379	10	AV187129	AV187129 AV187129
c	91	17	68.0	379	11	C70151	C70151 C70151 Yuji
c	92	17	68.0	385	11	T14661	T14661 05c04a07-r2
	93	17	68.0	388	10	AI077089	AI077089 TENU3236
c	94	17	68.0	395	10	AI834428	AI834428 606068A02
c	95	17	68.0	395	10	AW944303	AW944303 SD03340.3
	96	17	68.0	412	10	AA274370	AA274370 TgESTzz25
	97	17	68.0	413	10	AW330122	AW330122 TENU4868
c	98	17	68.0	439	13	AQ846660	AQ846660 LMAJFV1_1
	99	17	68.0	442	11	BI353763	BI353763 GM25034.3
	100	17	68.0	445	13	TA167C01P	AL474827 T. brucei

ALIGNMENTS

```
RESULT   1
AV439422/c
LOCUS       AV439422       538 bp    mRNA            EST       23-AUG-2000
DEFINITION  AV439422 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PS050c10_r 5', mRNA sequence.
ACCESSION   AV439422
VERSION     AV439422.1  GI:8594647
KEYWORDS    EST.
SOURCE      Porphyra yezoensis.
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 538)
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .538
                     /organism="Porphyra yezoensis"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PS050c10_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      114 a    158 c    170 g     96 t
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 10;  Length 538;
  Best Local Similarity   87.5%;  Pred. No. 2.9e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcgat 25
           |||||||||| || ||||||| ||
Db     385 CAGGTGCATCCCCTGCATTTCCAT 362


RESULT   2
AV748431
LOCUS       AV748431       548 bp    mRNA            EST       19-OCT-2000
DEFINITION  AV748431 NPC Homo sapiens cDNA clone NPCAOF03 5', mRNA sequence.
ACCESSION   AV748431
VERSION     AV748431.1  GI:10906279
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   Song,H., Peng,Y., Gu,Y., Yang,Y., Gao,G., Xiao,H., Xu,X., Li,N.,
            Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu
            ,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Ye,M., Zhang,Q.
            , Han,Z., Chen,Z., Hu,R. and Chen,J.
  TITLE     Homo sapiens NPC library cDNA clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Qinghua Zhang
            Shanghai Institute of Endocrinology, Rui-Jin Hospital
            197 Rui-Jin II Road, Shanghai 200025, P. R. China
            Tel: 86-21-64370045(ex.663332)
            Fax: 86-21-64743206
            Email: mbshi@ms.stn.sh.cn
            This clone is available at Shanghai Hematology Institute in
            Shanghai.
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi- Tech Park, Pudong.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NPCAOF03"
                     /clone_lib="NPC"
                     /tissue_type="pituitary"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      147 a    146 c    131 g    119 t      5 others
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 10;  Length 548;
  Best Local Similarity   87.5%;  Pred. No. 2.9e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcgat 25
           |||| ||||| |||||||||| ||
Db     168 CAGGGGCATCCCCAGCATTTCTAT 191


RESULT   3
B07718
LOCUS       B07718         715 bp    DNA             GSS       15-NOV-1997
DEFINITION  736I1E0370318 Rhodobacter sphaeroides 2.4.1 genomic DNA library
            Rhodobacter sphaeroides genomic clone 736I1E0370318 similar to dfp
            (P44953), DNA sequence.
ACCESSION   B07718
VERSION     B07718.1  GI:2055510
KEYWORDS    GSS.
SOURCE      Rhodobacter sphaeroides.
  ORGANISM  Rhodobacter sphaeroides
            Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
            Rhodobacter.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Choudhary,M., Mackenzie,C., Nereng,K., Sodergren,E., Weinstock,G.M.
            and Kaplan,S.
  TITLE     Low-resolution sequencing of Rhodobacter sphaeroides 2.4.1T:
            chromosome II is a true chromosome
  JOURNAL   Microbiology 143, 3085-3099 (1997)
  MEDLINE   98015398
COMMENT     Contact: Mackenzie, C.
            Department of Microbiology and Molecular Genetics
            The University of Texas Health Science Center
            6431 Fannin St., Houston, TX 77006, USA
            Tel: (713) 794-1742
            Fax: (713) 794-1782
            Email: mackenz@utmmg.med.uth.tmc.edu
            Seq primer: pBluescript SK (-) T7
```

```
          Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Rhodobacter sphaeroides"
                     /strain="2.4.1T"
                     /db_xref="taxon:1063"
                     /clone="736I1E0370318"
                     /clone_lib="Rhodobacter sphaeroides 2.4.1 genomic DNA
                     library"
                     /lab_host="E. coli S17-1"
                     /note="Vector: pLA2917; Genomic DNA from Rhodobacter
                     sphaeroides was prepared and partially  digested with
                     Sau3AI. Size selected (20kb) fragments were subcloned into
                     the BglII site of the cosmid vector pLA2917 (Allen, L. N.
                     and R. S.  Hanson. 1985. Journal of Bacteriology,
                     161:955-962. The library was then  ordered around
                     chromosome II of Rhodobacter sphaeroides (Choudhary et
                     al., 1994. Journal of Bacteriology, 176:7694-7702). The
                     cosmids were then  digested with a variety of restriction
                     enzymes (BamHI, EcoRI, BglII,  PstI, EcoRV, NotI and
                     DNaseI) and restriction fragments subcloned into  the
                     respective multiple cloning site sites of  pBluescript SK
                     (-). Note  BglII fragments were subcloned into the
                     pBluescript BamHI site. DNaseI  fragments were subcloned
                     into the EcoRV site.  All subclones were  transformed into
                     E. coli XL1Blue MRF'. All fragments were then sequenced
                     and the sequences where possible were assembled using the
                     GCG program  GELASSEMBLE. "
BASE COUNT       97 a    238 c    255 g    107 t     18 others
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 13;  Length 715;
  Best Local Similarity   87.5%;  Pred. No. 3e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcga 24
          |||||||||||||  |||  |||||
Db     42 GCAGGTGCATCGCGAGCTCTTCGA 65


RESULT   4
AQ288588
LOCUS       AQ288588      801 bp    DNA             GSS       03-DEC-1998
DEFINITION  nbxb0033K04f CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0033K04f, DNA sequence.
ACCESSION   AQ288588
VERSION     AQ288588.1  GI:3950202
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 801)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence stop: 406.
FEATURES             Location/Qualifiers
     source          1. .801
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbxb0033K04f"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world  population especially those
                     inhabiting highly populated areas of the  humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a  haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991).  The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to  facilitate positional cloning, physical mapping
                     and genome sequencing  of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare  variety. The
                     library contains 36,864 clones with an average insert size
                     of 128.5 Kb providing 10.9 haploid genome equivalents. The
                     deep  coverage allows the isolation a particular sequence
                     with a probability  of 99.9 %. Two high density filters,
                     each containing 18,432 clones  (doubly spotted), represent
                     the whole library for colony screening."
BASE COUNT      213 a    214 c    162 g    212 t
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 13;  Length 801;
  Best Local Similarity   87.5%;  Pred. No. 3e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 caggtgcatcgccagcatttcgat 25
          | |||||||||||||  ||| |||||
Db    462 CCGGTGCATCGCCACCATGTCGAT 485


RESULT   5
AA879064/c
LOCUS       AA879064      376 bp    mRNA            EST       24-SEP-1998
DEFINITION  oj18a12.s1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1492510 3'
            similar to TR:Q14437 Q14437 GOLGI APPARATUS GP25L2 PROTEIN. ;, mRNA
            sequence.
ACCESSION   AA879064
VERSION     AA879064.1  GI:2988029
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 376)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 453   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 159.
FEATURES             Location/Qualifiers
     source          1. .376
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1492510"
```

```
                     /clone_lib="NCI_CGAP_Kid5"
                     /tissue_type="2 pooled tumors (clear cell type)"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAATATTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       91 a     93 c    100 g     92 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 10;  Length 376;
  Best Local Similarity   84.0%;  Pred. No. 5e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||| |||||| || || |||||||
Db     197 GCAGATGCATCACCTGCCTTTCGAT 173


RESULT   6
BG455205/c
LOCUS       BG455205      627 bp    mRNA            EST       19-MAR-2001
DEFINITION  NF102G09PL1F1070 Phosphate starved leaf Medicago truncatula cDNA
            clone NF102G09PL 5', mRNA sequence.
ACCESSION   BG455205
VERSION     BG455205.1  GI:13378530
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Liu,J., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores
            ,H.R., Inman,J.T., Weller,J.W., May,G.D. and Harrison,M.J.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula phosphate-starved leaf library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Harrison MJ
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7325
            Fax: 580 221 7380
            Email: mjharrison@noble.org
            Insert Length: 627   Std Error: 0.00
            Plate: 102  row: G  column: 09
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="Medicago truncatula"
                     /db_xref="taxon:3880"
                     /clone="NF102G09PL"
                     /clone_lib="Phosphate starved leaf"
                     /tissue_type="leaf"
                     /dev_stage="trifoliate"
                     /note="Vector: Lambda Zap; At the trifoliate stage, M.
                     truncatula plants were transplanted to phosphate-free sand
                     and grown for a further 30 days.  During this 30 day
                     period, the plants were fertilized twice weekly with 1/2
                     Hoaglands solution containing only 20uM potassium
                     phosphate.  RNA was prepared from above ground tissues."
BASE COUNT      147 a    141 c    167 g    172 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 11;  Length 627;
  Best Local Similarity   84.0%;  Pred. No. 5.3e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcgat 25
           |||| || |||||||||||| | |||
Db     375 GCAGCTGAATCGCCAGCATGTTGAT 351


RESULT   7
AI329089/c
LOCUS       AI329089      231 bp    mRNA            EST       28-DEC-1998
DEFINITION  a9g12ne.r1 Neurospora crassa evening cDNA library Neurospora crassa
            cDNA clone a9g12ne 3', mRNA sequence.
ACCESSION   AI329089
VERSION     AI329089.1  GI:4065648
KEYWORDS    EST.
SOURCE      Neurospora crassa.
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariales; Sordariaceae; Neurospora.
REFERENCE   1  (bases 1 to 231)
  AUTHORS   Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.
  TITLE     Two Neurospora crassa EST Databases
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: a9g12ne.f1
            Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
            Department of Chemistry and Biochemistry
            Advanced Center for Genome Technology, University of Oklahoma
            620 Parrington Oval, Norman, OK 73019, USA
            Tel: 405 325 4912
            Fax: 405 325 7762
            Email: broe@ou.edu
            We anticipate the future release of the cDNA clones to the Fungal
            Genetics Stock Center
            Seq primer: Universal Reverse Primer
            High quality sequence stop: 179.
FEATURES             Location/Qualifiers
     source          1. .231
                     /organism="Neurospora crassa"
                     /strain="Strain 30-7 (bd; A)"
                     /db_xref="taxon:5141"
                     /clone="a9g12ne"
                     /clone_lib="Neurospora crassa evening cDNA library"
                     /tissue_type="tissue harvested following 22hr growth in
                     dark"
                     /note="Vector: pBlueScript SK-; Site_1: XbaI; Site_2:
                     EcoRI; See: Bell-Perdersen,D., et al. PNAS 93:13096,1996.
                     5' end of cDNA cloned into XbaI site of pBluescript; 3'
                     end of cDNA cloned into EcoRI site of pBluescript"
BASE COUNT       50 a     71 c     39 g     71 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 231;
  Best Local Similarity   87.0%;  Pred. No. 6.9e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 aggtgcatcgccagcatttcgat 25
           ||| |||||||||||||||  |||
Db      61 AGGGGCATCGCCAGCATTGGGAT 39


RESULT   8
AQ849360
LOCUS       AQ849360      403 bp    DNA             GSS       25-MAY-2001
DEFINITION  LMAJFV1_lm45g09.y1 Leishmania major FV1 random genomic library
            Leishmania major genomic clone LMAJFV1_lm45g09 5' similar to
            TR:O15206 O15206 AXONEMAL DYNEIN HEAVY CHAIN ;, DNA sequence.
ACCESSION   AQ849360
VERSION     AQ849360.1  GI:6054008
KEYWORDS    GSS.
SOURCE      Leishmania major.
```

```
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 403)
  AUTHORS   Akopyants,N.S., Clifton,S.W., Martin,J., Pape,D., Wylie,T., Li,L.,
            Kissinger,J., Roos,D.S., Marra,M., Hillier,L., Chinwalla,A.,
            Blistain,A., Schmitt,A., Person,B., Theising,B., Ritter,E., Ronko
            ,I., Bennett,J., Cole,R., Underwood,K., Cardenas,M., Gibbons,M.,
            Harvey,N., McCann,R., Tsaqareishvili,R., Williams,T., Jackson,Y.,
            Bowers,Y., Swaller,T., Waterston,R., Wilson,R. and Beverley,S.M.
  TITLE     A survey of the Leishmania major Friedlin strain V1 genome by
            shotgun sequencing: a resource for DNA microarrays and expression
            profiling
  JOURNAL   Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
  MEDLINE   21192569
COMMENT     Contact: Akopyants, NS / Beverley, SM
            WashU Leishmania Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library construction: Natalia S. Akopyants, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             If using this information please cite:
             N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
            Friedlin strain V1 genome by shotgun sequencing' and the Washington
            University Genome Sequencing Center For information on obtaining
            clone material please contact: Natalia S. Akopyants Ph.D.
            (natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
            (beverley@borcim.wustl.edu)
            Trace considered overall poor quality
            Seq primer: -40RP from Gibco
            Class: shotgun
            High quality sequence stop: 1.
FEATURES              Location/Qualifiers
     source           1. .403
                      /organism="Leishmania major"
                      /strain="Friedlin strain V1"
                      /db_xref="taxon:5664"
                      /clone="LMAJFV1_1m45g09"
                      /clone_lib="Leishmania major FV1 random genomic library"
                      /lab_host="TOP10 (Invitrogen)"
                      /note="Vector: pZero-2 (Invitrogen); Site_1: EcoRV;
                      Genomic DNA was isolated from stationary phase cells. For
                      this library, DNA was sheared to give a tight size
                      distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
                      polymerase, dephosphorylated with Shrimp Alkaline
                      Phosphatase and ligated into pZero-2 vector's EcoRV site."
BASE COUNT       75 a    121 c    131 g     76 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 13;  Length 403;
  Best Local Similarity   87.0%;  Pred. No. 7.4e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 aggtgcatcgccagcatttcgat 25
           |||| ||| ||| ||||||||||
Db     141 AGGTCCATGGCCGGCATTTCGAT 163


RESULT  9
P977L
LOCUS       P977L         527 bp    DNA             GSS       25-JUL-2000
DEFINITION  Leishmania major Friedlin PAC P977 left end-sequence, similar to
            TR:Q9VH71 Q9VH71 CG12946 PROTEIN. [0] +2. . ., N=56, Prob=8.8e-14,
            genomic survey sequence.
ACCESSION   AL390686
VERSION     AL390686.1  GI:9501662
KEYWORDS    GSS.
SOURCE      Leishmania major.
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and
            Smith,D.F.
  TITLE     A physical map of the Leishmania major Friedlin genome
  JOURNAL   Genome Res. 8 (2), 135-145 (1998)
  MEDLINE   98146435
REFERENCE   2  (bases 1 to 527)
  AUTHORS   Taylor,R.G., Huckle,E.E.J., Ivens,A.C., Rajandream,M.A. and
            Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            alicat@sanger.ac.uk
COMMENT     see http://www.ebi.ac.uk/parasites/leish.html
            Details of Leishmania sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/L_major/
            The primer sequence can be obtained from alicat@sanger.ac.uk.
FEATURES              Location/Qualifiers
     source           1. .527
                      /organism="Leishmania major"
                      /strain="Friedlin"
                      /db_xref="taxon:5664"
                      /clone="PAC P977"
BASE COUNT       76 a    197 c    169 g     85 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 13;  Length 527;
  Best Local Similarity   87.0%;  Pred. No. 7.6e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcg 23
           |||||||| |||||||||  |||
Db     282 GCAGGTGCGTCGCCAGCACGTCG 304


RESULT  10
BI235362/c
LOCUS       BI235362      594 bp    mRNA            EST       12-JUL-2001
DEFINITION  RE31201.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE31201 5 similar to CG7334:
            FBan0007334 'transporter' located on: 3L 68C13-68C13;: 05/12/2001,
            mRNA sequence.
ACCESSION   BI235362
VERSION     BI235362.1  GI:14703805
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003544: arm:3L [11370031,11651899]
            estimated-cyto:68C3-68C15: 05/12/2001
            Plate: RE.312  row: A  column: 1
            High quality sequence stop: 543.
```

```
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RE31201"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT      141 a    170 c    140 g    143 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 11;  Length 594;
  Best Local Similarity   87.0%;  Pred. No. 7.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 aggtgcatcgccagcatttcgat 25
          ||| || |||||||||||| ||||
Db    552 AGGGGCTTCGCCAGCATTCCGAT 530


RESULT  11
BI228734/c
LOCUS       BI228734      597 bp    mRNA            EST       11-JUL-2001
DEFINITION  RE26369.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE26369 5 similar to CG7334:
            FBan0007334 'transporter' located on: 3L 68C13-68C13;: 04/12/2001,
            mRNA sequence.
ACCESSION   BI228734
VERSION     BI228734.1  GI:14695998
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003544: arm:3L [11370031,11651899]
            estimated-cyto:68C3-68C15: 04/12/2001
            Plate: RE.263  row: F  column: 9
            High quality sequence stop: 587.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RE26369"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT      140 a    172 c    140 g    145 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 11;  Length 597;
  Best Local Similarity   87.0%;  Pred. No. 7.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 aggtgcatcgccagcatttcgat 25
          ||| || |||||||||||| ||||
Db    552 AGGGGCTTCGCCAGCATTCCGAT 530


RESULT  12
BI372325/c
LOCUS       BI372325      630 bp    mRNA            EST       01-AUG-2001
DEFINITION  RE59263.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE59263 5 similar to CG7334:
            FBan0007334 GO:[transporter (GO:0005215)] located on: 3L
            68C13-68C13;: 05/16/2001, mRNA sequence.
ACCESSION   BI372325
VERSION     BI372325.1  GI:15068353
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 630)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003544: arm:3L [11370031,11651899]
            estimated-cyto:68C3-68C15: 05/16/2001
            Plate: RE.592  row: F  column: 3
            High quality sequence stop: 568.
FEATURES             Location/Qualifiers
     source          1. .630
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RE59263"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT      146 a    181 c    150 g    153 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 11;  Length 630;
  Best Local Similarity   87.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 aggtgcatcgccagcatttcgat 25
          ||| || |||||||||||| ||||
```

```
Db    557 AGGGGCTTCGCCAGCATTCCGAT 535


RESULT 13
BI228776/c
LOCUS       BI228776      659 bp    mRNA            EST       11-JUL-2001
DEFINITION  RE26419.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE26419 5 similar to CG7334:
            FBan0007334 'transporter' located on: 3L 68C13-68C13;: 04/12/2001,
            mRNA sequence.
ACCESSION   BI228776
VERSION     BI228776.1  GI:14696040
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003544: arm:3L [11370031,11651899]
            estimated-cyto:68C3-68C15: 04/12/2001
            Plate: RE.264  row: B  column: 7
            High quality sequence stop: 603.
FEATURES             Location/Qualifiers
     source          1. .659
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RE26419"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT      147 a    190 c    159 g    162 t      1 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 11;  Length 659;
  Best Local Similarity   87.0%;  Pred. No. 7.8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 aggtgcatcgccagcatttcgat 25
          ||| || |||||||||||| ||||
Db    552 AGGGGCTTCGCCAGCATTCCGAT 530


RESULT 14
AZ176195/c
LOCUS       AZ176195      834 bp    DNA             GSS       30-AUG-2000
DEFINITION  SP_0141_A1_A06_T7A Strongylocentrotus purpuratus, purple sea urchin
            , sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=141 Col=11 Row=A, DNA sequence.
ACCESSION   AZ176195
VERSION     AZ176195.1  GI:8348570
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 834)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 141  row: A  column: 11
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 834.
FEATURES             Location/Qualifiers
     source          1. .834
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="Plate=141 Col=11 Row=A"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
BASE COUNT      270 a    132 c    162 g    270 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 13;  Length 834;
  Best Local Similarity   87.0%;  Pred. No. 8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 caggtgcatcgccagcatttcga 24
          || ||||| |||||||||||| |
Db    368 CAAGTGCAACGCCAGCATTTCAA 346


RESULT 15
P858R
LOCUS       P858R         360 bp    DNA             GSS       25-JUL-2000
DEFINITION  Leishmania major Friedlin PAC P858 right end-sequence, similar to
            SW:PVDA_BURCE O51940 L-ORNITHINE 5-MONOOXYGENASE (EC. . ., N=107,
            Prob=1.9e-17, genomic survey sequence.
ACCESSION   AL390537
VERSION     AL390537.1  GI:9501513
KEYWORDS    GSS.
SOURCE      Leishmania major.
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and
            Smith,D.F.
  TITLE     A physical map of the Leishmania major Friedlin genome
  JOURNAL   Genome Res. 8 (2), 135-145 (1998)
  MEDLINE   98146435
REFERENCE   2  (bases 1 to 360)
  AUTHORS   Taylor,R.G., Huckle,E.E.J., Ivens,A.C., Rajandream,M.A. and
            Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
```

```
            alicat@sanger.ac.uk
COMMENT     see http://www.ebi.ac.uk/parasites/leish.html
            Details of Leishmania sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/L_major/
            The primer sequence can be obtained from alicat@sanger.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Leishmania major"
                     /strain="Friedlin"
                     /db_xref="taxon:5664"
                     /clone="PAC P858"
BASE COUNT       48 a    117 c    112 g     83 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 13;  Length 360;
  Best Local Similarity   90.5%;  Pred. No. 1.1e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 gtgcatcgccagcatttcgat 25
           ||||| |||||||| ||||||
Db     150 GTGCAGCGCCAGCAGTTCGAT 170


RESULT  16
AV186249/c
LOCUS       AV186249      384 bp    mRNA            EST       22-JUL-1999
DEFINITION  AV186249 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk497c6 5', mRNA sequence.
ACCESSION   AV186249
VERSION     AV186249.1  GI:5568232
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 384)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
            Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A., Sano
            ,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
            Nomoto,H.
  TITLE     Expressed genes in C.elegans
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yuji Kohara
            Genome Biology Lab.
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .384
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk497c6"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT      120 a     85 c     78 g    101 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 10;  Length 384;
  Best Local Similarity   90.5%;  Pred. No. 1.1e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 ggtgcatcgccagcatttcga 24
           | |||||| ||||||||||||
Db     292 GTTGCATCCCCAGCATTTCGA 272


RESULT  17
AQ686803/c
LOCUS       AQ686803      439 bp    DNA             GSS       01-JUL-1999
DEFINITION  nbxb0073I05r CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0073I05r, DNA sequence.
ACCESSION   AQ686803
VERSION     AQ686803.1  GI:5327971
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GGAAACAGCTATGACCATG
            Class: BAC ends
            High quality sequence stop: 323.
FEATURES             Location/Qualifiers
     source          1. .439
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbxb0073I05r"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world  population especially those
                     inhabiting highly populated areas of the  humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a  haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991).  The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to  facilitate positional cloning, physical mapping
                     and genome sequencing  of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare  variety. The
                     library contains 36,864 clones with an average insert size
                     of 128.5 Kb providing 10.9 haploid genome equivalents. The
                     deep  coverage allows the isolation a particular sequence
                     with a probability  of 99.9 %. Two high density filters,
                     each containing 18,432 clones  (doubly spotted), represent
                     the whole library for colony screening."
BASE COUNT      146 a     88 c     82 g    123 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 13;  Length 439;
  Best Local Similarity   90.5%;  Pred. No. 1.1e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 ggtgcatcgccagcatttcga 24
           |||||||||||| | ||||||
Db     111 GGTGCATCGCCAACCTTTCGA 91


RESULT  18
AI532390/c
LOCUS       AI532390      498 bp    mRNA            EST       23-APR-2001
DEFINITION  SD03891.5prime SD Drosophila melanogaster Schneider L2 cell culture
```

```
            pOT2 Drosophila melanogaster cDNA clone SD03891 5 similar to
            CG8473: FBan0008473 'transporter' located on: X 13E18-13F1;:
            04/13/2001, mRNA sequence.
ACCESSION   AI532390
VERSION     AI532390.1  GI:4446525
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: SD03891.3prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003500: arm:X [15448919,15776364]
            estimated-cyto:13E14-14A4: 04/13/2001
            Plate: SD.38  row: H  column: 7
            High quality sequence stop: 457
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .498
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="SD03891"
                      /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                      culture pOT2"
                      /lab_host="DH5-alpha"
                      /note="Vector: pOT2; Site_1: EcoR1; Site_2: Xho1; Sized
                      fractionated cDNAs were directly ligated into pOT2.
                      Plasmid cDNA library."
BASE COUNT      158 a    106 c    124 g    110 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 10;  Length 498;
  Best Local Similarity   90.5%;  Pred. No. 1.1e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 aggtgcatcgccagcatttcg 23
          | | ||||||||||||||||||
Db    270 ATGCGCATCGCCAGCATTTCG 250


RESULT  19
BI375275/c
LOCUS       BI375275       551 bp    mRNA            EST       01-AUG-2001
DEFINITION  RE63047.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE63047 5 similar to CG7334:
            FBan0007334 GO:[transporter (GO:0005215)] located on: 3L
            68C13-68C13;: 05/16/2001, mRNA sequence.
ACCESSION   BI375275
VERSION     BI375275.1  GI:15071303
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 551)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003544: arm:3L [11370031,11651899]
            estimated-cyto:68C3-68C15: 05/16/2001
            Plate: RE.630  row: D  column: 11
            High quality sequence stop: 467.
FEATURES              Location/Qualifiers
     source           1. .551
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="RE63047"
                      /clone_lib="RE Drosophila melanogaster normalized Embryo
                      pFlc-1"
                      /sex="male and female"
                      /dev_stage="0-24 hours mixed stage embryonic"
                      /lab_host="DH5-alpha TonA"
                      /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                      BamHI; Library was kindly generated by Piero Carninci at
                      the RIKEN. The library was normalized and excised using
                      Cre recombinase. Plasmid cDNA library."
BASE COUNT      133 a    156 c    129 g    132 t      1 others
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 11;  Length 551;
  Best Local Similarity   86.4%;  Pred. No. 1.1e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 ggtgcatcgccagcatttcgat 25
          || || ||||||||||| ||||
Db    551 GGGGCTTCGCCAGCATTNCGAT 530


RESULT  20
AQ291570/c
LOCUS       AQ291570       580 bp    DNA             GSS       03-DEC-1998
DEFINITION  nbxb0039K21f CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0039K21f, DNA sequence.
ACCESSION   AQ291570
VERSION     AQ291570.1  GI:3952764
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence stop: 437.
FEATURES              Location/Qualifiers
     source           1. .580
                      /organism="Oryza sativa"
                      /strain="Japonica"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:4530"
```

/clone="nbxb0039K21f"
/clone_lib="CUGI Rice BAC Library"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 194 a 111 c 105 g 168 t 2 others
ORIGIN

Query Match 71.2%; Score 17.8; DB 13; Length 580;
Best Local Similarity 90.5%; Pred. No. 1.1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 ggtgcatcgccagcatttcga 24
||||||||||||| | ||||||
Db 112 GGTGCATCGCCAACCTTTCGA 92

RESULT 21
BF203823
LOCUS BF203823 693 bp mRNA EST 06-NOV-2000
DEFINITION 601868810F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4111418 5', mRNA sequence.
ACCESSION BF203823
VERSION BF203823.1 GI:11097409
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 693)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM999 row: c column: 03
High quality sequence start: 48
High quality sequence stop: 54.
FEATURES Location/Qualifiers
source 1. .693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4111418"
/clone_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 126 a 191 c 202 g 174 t
ORIGIN

Query Match 71.2%; Score 17.8; DB 11; Length 693;
Best Local Similarity 90.5%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 ggtgcatcgccagcatttcga 24
|||||||| ||||| ||||||
Db 606 GGTGCATCCCCAGCTTTTCGA 626

RESULT 22
BG718987
LOCUS BG718987 733 bp mRNA EST 08-MAY-2001
DEFINITION 602699291F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4831127 5', mRNA sequence.
ACCESSION BG718987
VERSION BG718987.1 GI:13998174
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 733)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM10753 row: f column: 24
High quality sequence stop: 729.
FEATURES Location/Qualifiers
source 1. .733
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4831127"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 238 a 161 c 152 g 182 t
ORIGIN

Query Match 71.2%; Score 17.8; DB 11; Length 733;
Best Local Similarity 90.5%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 gtgcatcgccagcatttcgat 25
||||||| ||||||||| |||

```
Db    359 GTGCATCCCCAGCATTTGGAT 379


RESULT  23
BI160225/c
LOCUS       BI160225      887 bp    mRNA            EST       05-JUL-2001
DEFINITION  602863918F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:5018044 5',
            mRNA sequence.
ACCESSION   BI160225
VERSION     BI160225.1  GI:14620226
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 887)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1831  row: c  column: 05
            High quality sequence start: 6
            High quality sequence stop: 839.
FEATURES             Location/Qualifiers
     source          1. .887
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5018044"
                     /clone_lib="NIH_MGC_42"
                     /tissue_type="epithelioid carcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by Ling
                     Hong in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"
BASE COUNT      251 a    240 c    209 g    187 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 11;  Length 887;
  Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 caggtgcatcgccagcatttc 22
          |||| |||| ||||||||||
Db    712 CAGGGGCATAGCCAGCATTTC 692


RESULT  24
CNS003ZE
LOCUS       CNS003ZE      888 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR09N17 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065360
VERSION     AL065360.1  GI:4942712
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 888)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .888
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR09N17"
                     /note="end : T7"
BASE COUNT      201 a    219 c    215 g    204 t     49 others
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 13;  Length 888;
  Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcattt 21
          ||||||| |||||||||| |
Db    331 GCAGGTGCCTCGCCAGCATGT 351


RESULT  25
CNS058GS/c
LOCUS       CNS058GS      1090 bp    DNA             GSS       26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            026L08 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL325909
VERSION     AL325909.1  GI:8219498
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1090)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1090)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1090)
```

```
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1090
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="026L08"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA026DF04A1-end : T3"
BASE COUNT      216 a    269 c    338 g    249 t     18 others
ORIGIN


  Query Match           71.2%;  Score 17.8;  DB 13;  Length 1090;
  Best Local Similarity  90.5%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 gtgcatcgccagcatttcgat 25
           |||||||| ||||||||| ||
Db     989 GTGCATCGACAGCATTTCCAT 969


RESULT  26
AK018497/c
LOCUS       AK018497     1862 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus adult male colon cDNA, RIKEN full-length enriched
            library, clone:9030409G11, full insert sequence.
ACCESSION   AK018497
VERSION     AK018497.1  GI:12858226
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male colon cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:9030409G11.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1862)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1862)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1862)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1862)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)

REFERENCE   5  (bases 1 to 1862)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 10.0 and subtraction to
            Rot = 185.2. Second strand cDNA was prepared with the primer
            adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
            bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
            end: BamHI. Host: DH10B.
FEATURES             Location/Qualifiers
     source          1. .1862
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1904075"
                     /db_xref="MGD:MGI:1918779"
                     /clone="9030409G11"
                     /sex="male"
                     /tissue_type="colon"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
BASE COUNT      390 a    540 c    564 g    367 t      1 others
ORIGIN


  Query Match           71.2%;  Score 17.8;  DB 12;  Length 1862;
  Best Local Similarity  90.5%;  Pred. No. 1.3e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttc 22
           ||||| | ||||||||||||||
Db     478 CAGGTCCTTCGCCAGCATTTC 458


RESULT  27
BE089771
LOCUS       BE089771     317 bp    mRNA            EST       12-JUN-2000
DEFINITION  QV0-BT0704-120500-224-g09 BT0704 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE089771
VERSION     BE089771.1  GI:8480206
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=&t2=QV0-BT0704-120
            500-224-g09&t3=2000-05-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 20
            High quality sequence stop: 317.
FEATURES             Location/Qualifiers
     source          1. .317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0704"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       99 a     77 c     78 g     63 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 317;
  Best Local Similarity   83.3%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcga 24
          ||||||| ||| |||| | |||||
Db     59 GCAGGTGGATCACCAGGAGTTCGA 82


RESULT  28
BI371476
LOCUS       BI371476      411 bp    mRNA            EST       01-AUG-2001
DEFINITION  RE58374.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE58374 5 similar to Mkrn1:
            FBan0007184 GO:[] located on: 3L 78E1-78E1;: 05/16/2001, mRNA
            sequence.
ACCESSION   BI371476
VERSION     BI371476.1  GI:15067504
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003594: arm:3L [21112610,21418125]
            estimated-cyto:78C7-78E3: 05/16/2001
            Plate: RE.583  row: G  column: 2
            High quality sequence stop: 259.
FEATURES             Location/Qualifiers
     source          1. .411
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RE58374"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT      101 a    121 c    107 g     82 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 411;
  Best Local Similarity   83.3%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcga 24
          ||||  |||||||||||||| || |
Db    385 GCAGCCGCATCGCCAGCATCTCCA 408


RESULT  29
AQ890634/c
LOCUS       AQ890634      418 bp    DNA             GSS       10-NOV-1999
DEFINITION  HS_2204_A1_G03_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2204 Col=5 Row=M, DNA sequence.
ACCESSION   AQ890634
VERSION     AQ890634.1  GI:6346824
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2204  row: M  column: 5
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 418.
```

```
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2204 Col=5 Row=M"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      119 a     98 c    102 g     99 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 418;
  Best Local Similarity   83.3%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 caggtgcatcgccagcatttcgat 25
          ||  ||| ||||||||||||| ||
Db    144 CAACTGCTTCGCCAGCATTTCCAT 121


RESULT  30
AQ516531/c
LOCUS       AQ516531      450 bp    DNA             GSS       05-MAY-1999
DEFINITION  HS_5052_B1_D03_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=628 Col=5 Row=H, DNA sequence.
ACCESSION   AQ516531
VERSION     AQ516531.1  GI:4748679
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 628  row: H  column: 5
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 450.
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=628 Col=5 Row=H"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      107 a    105 c     94 g    142 t      2 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 450;
  Best Local Similarity   83.3%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcaggtgcatcgccagcatttcga 24
          ||||| | ||| |||| |||||||
Db    251 GCAGGCGGATCACCAGGATTTCGA 228


RESULT  31
BF654469/c
LOCUS       BF654469      475 bp    mRNA            EST       25-APR-2001
DEFINITION  278665 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF654469
VERSION     BF654469.1  GI:11919601
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 475)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 71  row: H  column: 16
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .475
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 3BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT       91 a    156 c    134 g     94 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 475;
  Best Local Similarity   83.3%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 caggtgcatcgccagcatttcgat 25
          |||||||||||| |  ||||| |||
Db    474 CAGGTGCATCGGCTTCATTTAGAT 451


RESULT  32
BG266805/c
```

```
LOCUS       BG266805      506 bp    mRNA            EST       20-FEB-2001
DEFINITION  1000101G12.x1 1000 - Unigene I from Maize Genome Project Zea mays
            cDNA, mRNA sequence.
ACCESSION   BG266805
VERSION     BG266805.1  GI:12970082
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 506)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 1000101  row: G  column: 12.
FEATURES             Location/Qualifiers
     source          1. .506
                     /organism="Zea mays"
                     /db_xref="dbEST:605008C11.x1"
                     /db_xref="taxon:4577"
                     /clone_lib="1000 - Unigene I from Maize Genome Project"
                     /note="This library represents the unique ESTs found in
                     the first round of EST sequencing at Stanford University
                     for the maize genome project. Sequences are present from
                     libraries 486, 487, 496, 603, 605, 606, 614, 618, 660, 683
                     , 687, 707, and 945. Contigs were assembled using TIGR's
                     CAP program and a representative EST from each contig was
                     selected for the Unigene set. All singlets were also
                     selected."
BASE COUNT       96 a    168 c    129 g    113 t
ORIGIN
```

```
  Query Match             70.4%;  Score 17.6;  DB 11;  Length 506;
  Best Local Similarity   83.3%;  Pred. No. 1.4e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcga 24
           |||||||||||  ||||||  | | |
Db     456 GCAGGTGCATCACCAGCAGTACCA 433
```

```
RESULT  33
BI423547
LOCUS       BI423547      518 bp    mRNA            EST       16-AUG-2001
DEFINITION  949049D11.y1 949 - Juvenile leaf and shoot cDNA from Steve Moose
            Zea mays cDNA, mRNA sequence.
ACCESSION   BI423547
VERSION     BI423547.1  GI:15198896
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 518)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 949049  row: D  column: 11.
FEATURES             Location/Qualifiers
     source          1. .518
                     /organism="Zea mays"
                     /cultivar="W64A"
                     /db_xref="taxon:4577"
                     /clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
                     Moose"
                     /tissue_type="immature leaf primordium and vegetative
                     meristem"
                     /dev_stage="4 stages from 3-13 days after imbibing"
                     /lab_host="E. coli XLOLR"
                     /note="Organ: juvenile vegetative shoots; Vector:
                     pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts
                     of total RNA by weight from 4 tissue sources (see below)
                     were pooled, polyA+ RNA isolated, and cDNA synthesized for
                     EcoRI (5') and XhoI (3') directional cloning into lambda
                     Hybrizap vector from Stratagene. Tissue Sources: 1. Whole
                     shoots 3 days after sowing/imbibing in wet soil. 2. Basal
                     1.5 cm shoots 6 days after sowing - includes yellow
                     portions of developing leaves 1-5, primordia from 6-8, and
                     the vegetative apex. 3. Non-green portions of developing
                     leaves 4-5 and the vegetative apex,  including adult leaf
                     primordia, 9 days after sowing. 4. Partially expanded and
                     greening leaves 4-5 at 13 days after sowing."
BASE COUNT      108 a    175 c    165 g     70 t
ORIGIN
```

```
  Query Match             70.4%;  Score 17.6;  DB 11;  Length 518;
  Best Local Similarity   83.3%;  Pred. No. 1.4e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcga 24
           |||||||||||  ||||||  | | |
Db     239 GCAGGTGCATCACCAGCAGTACCA 262
```

```
RESULT  34
AV619677
LOCUS       AV619677      549 bp    mRNA            EST       15-DEC-2000
DEFINITION  AV619677 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
            reinhardtii cDNA clone LC011f07_r 5', mRNA sequence.
ACCESSION   AV619677
VERSION     AV619677.1  GI:10768852
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="LC011f07_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
```

```
                    XhoI; The cDNA library was constructed from cells cultured
                    in a carbon stress acclimatized condition in which carbon
                    dioxide concentration in the bubbling gas was changed from
                    5% to 0.04%"
BASE COUNT      105 a    171 c    179 g     94 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 549;
  Best Local Similarity   83.3%;  Pred. No. 1.4e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcga 24
           ||||| ||| |||||||| ||| |
Db     338 GCAGGAGCAGCGCCAGCACTTCAA 361


RESULT  35
AQ795581
LOCUS       AQ795581      558 bp    DNA             GSS       04-AUG-1999
DEFINITION  nbxb0056A12f CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0056A12f, DNA sequence.
ACCESSION   AQ795581
VERSION     AQ795581.1  GI:5703966
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence stop: 334.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbxb0056A12f"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world  population especially those
                     inhabiting highly populated areas of the  humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a  haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991).  The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to  facilitate positional cloning, physical mapping
                     and genome sequencing  of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare  variety. The
                     library contains 36,864 clones with an average insert size
                     of 128.5 Kb providing 10.9 haploid genome equivalents. The
                     deep  coverage allows the isolation a particular sequence
                     with a probability  of 99.9 %. Two high density filters,
                     each containing 18,432 clones  (doubly spotted), represent
                     the whole library for colony screening."
BASE COUNT      122 a    177 c    139 g    120 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 558;
  Best Local Similarity   83.3%;  Pred. No. 1.4e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcgat 25
           | |||||||||||| ||  |||||
Db     462 CCGGTGCATCGCCACCACGTCGAT 485


RESULT  36
AI665214/c
LOCUS       AI665214      582 bp    mRNA            EST       02-FEB-2000
DEFINITION  605008C11.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
            cDNA, mRNA sequence.
ACCESSION   AI665214
VERSION     AI665214.2  GI:4804415
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     On May 11, 1999 this sequence version replaced gi:4776211.
            Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 605008  row: C  column: 11.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="605 - Endosperm cDNA library from Schmidt lab"
                     /tissue_type="nucellar, embryo, and endosperm"
                     /dev_stage="10-14 days post-pollination"
                     /lab_host="DH5(alpha)"
                     /note="Organ: Kernel; Vector: pAD-GAL4-2'; Site_1: EcoRI;
                     Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
                     lab"
BASE COUNT      115 a    191 c    146 g    130 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 582;
  Best Local Similarity   83.3%;  Pred. No. 1.4e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcga 24
           |||||||||| |||||| | | |
Db     483 GCAGGTGCATCACCAGCAGTACCA 460


RESULT  37
BF425225/c
LOCUS       BF425225      615 bp    mRNA            EST       28-NOV-2000
DEFINITION  su54h10.y1 Gm-c1069 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1069-1123 5' similar to SW:UBC9_ARATH P35132
            UBIQUITIN-CONJUGATING ENZYME E2-17 KD 9 ;, mRNA sequence.
ACCESSION   BF425225
VERSION     BF425225.1  GI:11413214
```

```
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 615)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            High quality sequence stop: 412.
FEATURES             Location/Qualifiers
     source          1. .615
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1069-1123"
                     /clone_lib="Gm-c1069"
                     /tissue_type="Degenerating cotyledons, 9-10 day old
                     etiolated seedling"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from degenerating cotyledons of 9-10 day old etiolated
                     seedlings for the cultivar Williams. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a
                     poly(dT) sequence with a XhoI restriction site. EcoRI
                     adapters were ligated  to the blunt-ended cDNA fragments
                     followed by XhoI digestion. The cDNA  fragments were
                     directionally cloned into the EcoRI-XhoI restriction site
                     of  the pBluescript vector. The ligated cDNA fragments
                     were transformed into  DH10B host cells (GibcoBRL). This
                     library was constructed in the laboratory  of Dr. Randy
                     Shoemaker."
BASE COUNT      148 a    165 c    131 g    171 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 615;
  Best Local Similarity   83.3%;  Pred. No. 1.4e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 caggtgcatcgccagcatttcgat 25
          || ||||||| ||||||| |||
Db    532 CAAGTGCATCATCAGCATTTGGAT 509


RESULT  38
CNS07883/c
LOCUS       CNS07883      944 bp    DNA             GSS       08-JUL-2001
DEFINITION  T7 end of clone BB0AA018F09 of library BB0AA from strain CBS 4732
            of Pichia angusta, genomic survey sequence.
ACCESSION   AL433657
VERSION     AL433657.1  GI:12217071
KEYWORDS    GSS.
SOURCE      Pichia angusta.
  ORGANISM  Pichia angusta
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 944)
  AUTHORS   Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.
            and Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
            angusta
  JOURNAL   FEBS Lett. 487 (1), 76-81 (2000)
  MEDLINE   20584723
REFERENCE   2  (bases 1 to 944)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
REFERENCE   3  (bases 1 to 944)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .944
                     /organism="Pichia angusta"
                     /strain="CBS 4732"
                     /db_xref="taxon:4905"
                     /clone="BB0AA018F09"
                     /clone_lib="BB0AA"
                     /note="end : T7"
     misc_feature    <41. .>937
                     /note="similar to Saccharomyces cerevisiae ORF YOL021c [
                     DIS3 ; 3'->5' exoribonuclease required for 3' end
                     formation of 5.8S rRNA ]"
                     /evidence=not_experimental
BASE COUNT      218 a    235 c    289 g    201 t      1 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 944;
  Best Local Similarity   83.3%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 caggtgcatcgccagcatttcgat 25
          ||| |||||| |||||| |||||
Db    292 CAGTAGCATCGGCAGCATGTCGAT 269


RESULT  39
CNS03VDY/c
LOCUS       CNS03VDY      1032 bp    DNA             GSS       18-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            061N02 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL262303
VERSION     AL262303.1  GI:7983929
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
```

```
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1032)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1032)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1032)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1032
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="061N02"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG061DG01LP1-end : T7"
BASE COUNT      214 a    296 c    324 g    194 t      4 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 1032;
  Best Local Similarity   83.3%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcgat 25
           ||  |||||||||| ||| |||||
Db     518 CATCTGCATCGCCACCATCTCGAT 495


RESULT  40
CNS03B6D
LOCUS       CNS03B6D     1039 bp    DNA             GSS       15-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            011022 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL236110
VERSION     AL236110.1  GI:7895245
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1039)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1039)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1039)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1039
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="011022"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG011BH11LP1-end : T7"
BASE COUNT      248 a    222 c    236 g    329 t      4 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 1039;
  Best Local Similarity   83.3%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcatttcga 24
           ||| ||  |||| ||||||||||||
Db     254 GCATGTTAATCGTCAGCATTTCGA 277


RESULT  41
BG027526/c
LOCUS       BG027526     1077 bp    mRNA            EST       24-JAN-2001
DEFINITION  602295377F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4390003 5',
            mRNA sequence.
ACCESSION   BG027526
VERSION     BG027526.1  GI:12416239
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1077)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10078  row: j  column: 20
            High quality sequence stop: 620.
FEATURES             Location/Qualifiers
     source          1. .1077
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4390003"
                     /clone_lib="NIH_MGC_86"
                     /tissue_type="osteosarcoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bone; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.533 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      316 a    231 c    352 g    178 t
ORIGIN
```

```
  Query Match             70.4%;  Score 17.6;  DB 11;  Length 1077;
  Best Local Similarity   83.3%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatttcgat 25
           ||| |||||| |||||| ||| ||
Db     446 CAGCTGCATCTCCAGCAGTTCCAT 423


RESULT  42
BF933544/c
LOCUS       BF933544      376 bp    mRNA            EST       22-JAN-2001
DEFINITION  CM1-NT0270-211200-661-g03 NT0270 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF933544
VERSION     BF933544.1  GI:12350868
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 376)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM1&t2=CM1-NT0270-
            211200-661-g03&t3=2000-12-21&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 10
            High quality sequence stop: 376.
FEATURES             Location/Qualifiers
     source          1. .376
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0270"
                     /dev_stage="Adult"
                     /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       77 a    104 c     99 g     96 t
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 11;  Length 376;
  Best Local Similarity   94.7%;  Pred. No. 1.6e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 ggtgcatcgccagcatttc 22
           |||||||||||||||| ||
Db     329 GGTGCATCGCCAGCATCTC 311


RESULT  43
AQ171736
LOCUS       AQ171736      527 bp    DNA             GSS       17-OCT-1998
DEFINITION  HS_3172_B2_G07_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3172 Col=14 Row=N, DNA sequence.
ACCESSION   AQ171736
VERSION     AQ171736.1  GI:3569103
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3172  row: N  column: 14
            Class: BAC ends
            High quality sequence stop: 527.
FEATURES             Location/Qualifiers
     source          1. .527
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3172 Col=14 Row=N"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      148 a    125 c    135 g    117 t      2 others
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 13;  Length 527;
  Best Local Similarity   94.7%;  Pred. No. 1.7e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatt 20
           ||||||||||| ||||||||
Db     185 CAGGTGCATCCCCAGCATT 203


RESULT  44
AQ350006
LOCUS       AQ350006      634 bp    DNA             GSS       07-MAY-1999
DEFINITION  RPCI11-117D10.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-117D10,
            DNA sequence.
ACCESSION   AQ350006
VERSION     AQ350006.1  GI:4177341
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 634)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI11-117D10.TV
```

```
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .634
                     /organism="Homo sapiens"
                     /db_xref="GDB:7544625"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-117D10"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      158 a    143 c    195 g    138 t
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 13;  Length 634;
  Best Local Similarity   94.7%;  Pred. No. 1.7e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 caggtgcatcgccagcatt 20
           |||||||||| ||||||||
Db     179 CAGGTGCATCCCCAGCATT 197


RESULT  45
CNS03HMW/c
LOCUS       CNS03HMW      786 bp    DNA             GSS       17-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            027C16 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL244481
VERSION     AL244481.1  GI:7965493
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 786)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 786)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .786
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="027C16"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG027BB08LP1~end : T7"
BASE COUNT      236 a    188 c    199 g    160 t      3 others
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 13;  Length 786;
  Best Local Similarity   94.7%;  Pred. No. 1.7e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gcaggtgcatcgccagcat 19
           |||| ||||||||||||||
Db     495 GCAGATGCATCGCCAGCAT 477


Search completed: December 26, 2001, 12:11:50
Job time: 6472 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:57:33 ; Search time 1755.64 Seconds
(without alignments)
234.917 Million cell updates/sec

Title: US-09-396-196F-7
Perfect score: 25
Sequence: 1 gcatcgccagcatttcgatcctcgt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by anaiysis of the total score distribution.

SUMMARIES

%

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	6	AR029499	AR029499 Sequence
2	25	100.0	1041	6	AR034916	AR034916 Sequence
3	25	100.0	1084	6	A11530	A11530 BioB gene o
4	25	100.0	1121	6	E00893	E00893 Genomic DNA
5	25	100.0	5793	1	ECOBIO	J04423 E.coli 7,8-
6	25	100.0	5872	6	A38246	A38246 Sequence 1
7	25	100.0	5872	6	A38251	A38251 Sequence 6
8	25	100.0	5872	6	A93674	A93674 Sequence 1
9	25	100.0	5872	6	A93679	A93679 Sequence 6
10	25	100.0	5872	6	AR101809	AR101809 Sequence
11	25	100.0	5872	6	AR101810	AR101810 Sequence
12	25	100.0	11022	1	AE000180	AE000180 Escherich
13	25	100.0	13501	1	AE005258	AE005258 Escherich
14	25	100.0	297816	1	AP002553	AP002553 Escherich
15	24	96.0	5526	1	AF250776	AF250776 Unculture
16	20.4	81.6	10593	1	AE003860	AE003860 Xylella f
17	19.2	76.8	965	1	AF250770	AF250770 Unculture
18	19.2	76.8	4000	1	AF250768	AF250768 Unculture
c 19	19.2	76.8	68419	2	AC016014	AC016014 Homo sapi
20	19.2	76.8	116334	2	AC084172	AC084172 Homo sapi
c 21	19.2	76.8	186616	2	AC026294	AC026294 Homo sapi
22	19.2	76.8	189025	2	AC090154	AC090154 Homo sapi
c 23	18.6	74.4	80954	2	AC011276	AC011276 Homo sapi
c 24	18.6	74.4	112808	2	AC068024	AC068024 Homo sapi
c 25	18.6	74.4	159738	2	AC055855	AC055855 Homo sapi
26	18.2	72.8	1041	1	EHBIOTOP2	U38648 Erwinia her
c 27	18.2	72.8	3291	8	D88424	D88424 Cyanidium c
c 28	18.2	72.8	3451	6	E13998	E13998 Cyanidium c
c 29	18.2	72.8	4304	1	KPU31464	U31464 Klebsiella
30	18.2	72.8	10239	1	AE004457	AE004457 Pseudomon
31	18.2	72.8	11382	1	RLE310723	AJ310723 Rhizobium
c 32	18.2	72.8	11840	1	AE004541	AE004541 Pseudomon
33	18	72.0	128	1	ECOBIOB	M27731 Escherichia
34	17.8	71.2	3748	1	MSU31280	U31280 Methylobaci
35	17.8	71.2	3748	12	CVU39574	U39574 Expression
36	17.8	71.2	12107	1	AE002396	AE002396 Neisseria
c 37	17.8	71.2	14645	2	AC018306	AC018306 Drosophil
c 38	17.8	71.2	106462	2	AP003375	AP003375 Oryza sat
39	17.8	71.2	110000	2	TBBCHR1A_03	Continuation (4 of
40	17.8	71.2	166857	3	AC093192	AC093192 Drosophil
41	17.8	71.2	183439	3	AC007809	AC007809 Drosophil
c 42	17.8	71.2	195767	1	NMA7Z2491	AL162758 Neisseria
43	17.8	71.2	232744	3	AE003705	AE003705 Drosophil
c 44	17.8	71.2	237119	2	AC017740	AC017740 Drosophil
45	17.8	71.2	257867	3	AC005557	AC005557 Drosophil
46	17.8	71.2	303191	3	AE003472	AE003472 Drosophil
c 47	17.8	71.2	307551	1	AP003015	AP003015 Mesorhizo
c 48	17.8	71.2	329709	1	AP002997	AP002997 Mesorhizo
49	17.8	71.2	349980	6	AX044029	AX044029 Sequence
50	17.6	70.4	300	9	HSU10478	U10478 Human clone
51	17.6	70.4	327	1	AF227941	AF227941 Marine st
c 52	17.6	70.4	1653	3	AF034644	AF034644 Haematobi
53	17.6	70.4	7215	1	SMABIO	D17468 S. marcesce
54	17.6	70.4	11003	1	AF057718	AF057718 Agrobacte
c 55	17.6	70.4	11864	1	AE007930	AE007930 Agrobacte
56	17.6	70.4	12109	15	AB006858	Ab006858 Agrobacte
57	17.6	70.4	53425	1	EAU67194	U67194 Enterobacte
58	17.6	70.4	97900	3	AC087422	AC087422 Caenorhab
59	17.6	70.4	104287	9	AC005905	AC005905 Homo sapi
c 60	17.6	70.4	108845	1	U66917	U66917 Pseudomonas
61	17.6	70.4	110000	2	AL354832_1	Continuation (2 of
62	17.6	70.4	148018	2	AP003974	AP003974 Oryza sat
c 63	17.6	70.4	178062	9	CNS01DRK	AL117356 Human chr
c 64	17.6	70.4	206479	1	AB016260	AB016260 Agrobacte
65	17.6	70.4	222062	2	AC024149	AC024149 Homo sapi
66	17.4	69.6	459	8	MZESUSYCD	J01240 maize sucro
67	17.4	69.6	818	8	MZESUSYSH	J01241 maize sucro
68	17.4	69.6	2746	8	ZMSUCS1	X02400 Maize mRNA
69	17.4	69.6	6386	6	AR075162	AR075162 Sequence
70	17.4	69.6	6922	8	ZMSUCS2	X02382 Maize gene

c	71	17.4	69.6	112576	2	AC011250	AC011250 Drosophil
c	72	17.4	69.6	118591	2	AC020202	AC020202 Drosophil
c	73	17.4	69.6	119797	2	AC013236	AC013236 Drosophil
c	74	17.4	69.6	123647	3	AC007837	AC007837 Drosophil
c	75	17.4	69.6	128626	3	AC092245	AC092245 Drosophil
c	76	17.4	69.6	156753	3	AC009218	AC009218 Drosophil
c	77	17.4	69.6	260686	3	AE003791	AE003791 Drosophil
	78	17.4	69.6	264853	3	AE003644	AE003644 Drosophil
	79	17.4	69.6	300962	3	DROSADH04	AE003410 Drosophil
	80	17.2	68.8	1669	1	BBA224769	AJ224769 Bordetell
	81	17.2	68.8	1669	1	BPA224768	AJ224768 Bordetell
c	82	17.2	68.8	6722	5	AB030482	AB030482 Fugu pard
c	83	17.2	68.8	10683	1	AE008052	AE008052 Agrobacte
	84	17.2	68.8	10786	1	AE003906	AE003906 Xylella f
	85	17.2	68.8	15389	1	BPLPSBLOC	X90711 Bordetella
c	86	17.2	68.8	41642	1	BBR007747	AJ007747 Bordetell
c	87	17.2	68.8	42991	1	U00018	U00018 Mycobacteri
c	88	17.2	68.8	52158	2	AC013185	AC013185 Drosophil
	89	17.2	68.8	64884	8	NC65E11	AL451021 Neurospor
	90	17.2	68.8	67453	8	NCB12K8	AL513464 Neurospor
c	91	17.2	68.8	92569	2	AC016445	AC016445 Drosophil
	92	17.2	68.8	110000	2	LMFLCHR36_30	Continuation (31 o
c	93	17.2	68.8	142203	2	AC079521	AC079521 Mus muscu
c	94	17.2	68.8	157175	9	AL162713	AL162713 Human DNA
	95	17.2	68.8	177857	3	AC010211	AC010211 Drosophil
	96	17.2	68.8	207411	2	AC084746	AC084746 Mus muscu
c	97	17.2	68.8	266574	2	AC087129	AC087129 Mus muscu
c	98	17.2	68.8	291545	3	AE003600	AE003600 Drosophil
	99	17.2	68.8	306650	1	SME591783	AL591783 Sinorhizo
c	100	17.2	68.8	308950	1	MLEPRTN9	AL583925 Mycobacte

ALIGNMENTS

```
RESULT   1
AR029499
LOCUS       AR029499      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1  GI:5941472
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.Andrew.
  TITLE     Enhanced biotin biosynthesis in plant tissue
  JOURNAL   Patent: US 5859335-A 7 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.38;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcgt 25
          |||||||||||||||||||||||||
Db     90 GCATCGCCAGCATTTCGATCCTCGT 114


RESULT   2
AR034916
LOCUS       AR034916      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1  GI:5950521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.A.
  TITLE     Transgenic plants having increased biotin content
  JOURNAL   Patent: US 5869719-A 7 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.38;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcgt 25
          |||||||||||||||||||||||||
Db     90 GCATCGCCAGCATTTCGATCCTCGT 114


RESULT   3
A11530
LOCUS       A11530       1084 bp    DNA             PAT       09-FEB-1994
DEFINITION  BioB gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1  GI:490218
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   .
  JOURNAL   Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            24. .1064
                     /gene="BIOB"
     CDS             24. .1064
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490219"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
BASE COUNT      271 a    286 c    318 g    209 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1084;
  Best Local Similarity  100.0%;  Pred. No. 0.38;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcgt 25
          |||||||||||||||||||||||||
Db    113 GCATCGCCAGCATTTCGATCCTCGT 137


RESULT   4
E00893
LOCUS       E00893       1121 bp    DNA             PAT       29-SEP-1997
DEFINITION  Genomic DNA encoding biotin Synthetase.
```

```
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES             Location/Qualifiers
     source          1. .1121
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      289 a    296 c    325 g    211 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 0.38;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcgt 25
          |||||||||||||||||||||||||
Db    131 GCATCGCCAGCATTTCGATCCTCGT 155


RESULT   5
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES             Location/Qualifiers
     source          1. .5793
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     CDS             complement(98. .574)
                     /note="ORF 1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23513.1"
                     /db_xref="GI:455168"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene            complement(633. .1925)
                     /gene="bioA"
     CDS             complement(633. .1925)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase"
                     /protein_id="AAA23514.1"
                     /db_xref="GI:457106"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                     ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                     SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                     GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                     LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                     QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                     PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            2012. .3052
                     /gene="bioB"
     CDS             2012. .3052
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAA23515.1"
                     /db_xref="GI:145425"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3049. .4203
                     /gene="bioF"
     CDS             3049. .4203
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-amino pelargonic acid synthetase"
                     /protein_id="AAA23516.1"
                     /db_xref="GI:145426"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4190. .4945
                     /gene="bioC"
     CDS             4190. .4945
                     /gene="bioC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23517.1"
                     /db_xref="GI:145427"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
```

```
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4938. .5597
                     /gene="bioD"
     CDS             4938. .5597
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAA23518.1"
                     /db_xref="GI:145428"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                     T"
BASE COUNT     1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match           100.0%;  Score 25;  DB 1;  Length 5793;
  Best Local Similarity  100.0%;  Pred. No. 0.37;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcgt 25
          |||||||||||||||||||||||||
Db   2101 GCATCGCCAGCATTTCGATCCTCGT 2125


RESULT   6
A38246
LOCUS       A38246       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /number=1
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAA02325.1"
                     /db_xref="GI:2294845"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=3
                     /product="PROTEIN"
                     /protein_id="CAA02326.1"
                     /db_xref="GI:2294846"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:
                     8-AMINO-7-OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /number=5
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAA02327.1"
                     /db_xref="GI:2294847"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=6
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAA02328.1"
                     /db_xref="GI:2294848"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
```

```
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a    1552 c    1695 g    1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.37;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||||||||||||||||||||||||
Db     206 GCATCGCCAGCATTTCGATCCTCGT 230


RESULT   7
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLF1SGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a    1552 c    1695 g    1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.37;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||||||||||||||||||||||||
Db     206 GCATCGCCAGCATTTCGATCCTCGT 230


RESULT   8
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
```

```
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIER1DVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN
```

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||||||||||||||||||||||||
Db     206 GCATCGCCAGCATTTCGATCCTCGT 230
```

```
RESULT   9
A93679
LOCUS       A93679       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN
```

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.37;

```
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||||||||||||||||||||||||
Db     206 GCATCGCCAGCATTTCGATCCTCGT 230


RESULT  10
AR101809
LOCUS       AR101809     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 1 from patent US 6083712.
ACCESSION   AR101809
VERSION     AR101809.1  GI:12812607
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.37;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||||||||||||||||||||||||
Db     206 GCATCGCCAGCATTTCGATCCTCGT 230


RESULT  11
AR101810
LOCUS       AR101810     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 6 from patent US 6083712.
ACCESSION   AR101810
VERSION     AR101810.1  GI:12812608
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.37;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||||||||||||||||||||||||
Db     206 GCATCGCCAGCATTTCGATCCTCGT 230


RESULT  12
AE000180
LOCUS       AE000180    11022 bp    DNA             BCT       01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 70 of 400 of the complete
            genome.
ACCESSION   AE000180 U00096
VERSION     AE000180.1  GI:1786988
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 11022)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
  PUBMED    9278503
REFERENCE   2  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 11022)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data.  Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .11022
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   <1. .130
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 4 REP sequences"
     gene            complement(147. .1430)
                     /gene="ybhC"
                     /note="b0772"
```

```
CDS             complement(147. .1430)
                /gene="ybhC"
                /function="orf; Not classified"
                /note="f427; 98 pct identical to fragment YBHC_ECOLI
                SW:P46130 (300 aa) but contains 127 additional C-terminal
                residues"
                /codon_start=1
                /transl_table=11
                /product="putative pectinesterase"
                /protein_id="AAC73859.1"
                /db_xref="GI:1786989"
                /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPACTQGVTHTTIQAAVD
                AAIIKRTNKRQYIAVMPGEYQGTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
                NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSNIYYGFLAVNSRFN
                AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
promoter        complement(1507. .1535)
                /note="factor Sigma70; predicted +1 start at 806574"
gene            complement(1582. .2058)
                /gene="ybhB"
                /note="b0773"
CDS             complement(1582. .2058)
                /gene="ybhB"
                /function="orf; Unknown"
                /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC73860.1"
                /db_xref="GI:1786990"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
protein_bind    complement(2108. .2124)
                /note="central position to predicted promoter:85.5"
                /bound_moiety="RhaS predicted site"
gene            complement(2117. .3406)
                /gene="bioA"
                /note="b0774"
CDS             complement(2117. .3406)
                /gene="bioA"
                /EC_number="2.6.1.62"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                /codon_start=1
                /transl_table=11
                /product="7,8-diaminopelargonic acid synthetase"
                /protein_id="AAC73861.1"
                /db_xref="GI:1786991"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
promoter        complement(2193. .2221)
                /gene="bioA"
                /note="factor Sigma70; predicted +1 start at 807260"
protein_bind    3411. .3450
                /note="central position to bioB promoter: -20"
                /bound_moiety="BioB documented site"
protein_bind    3411. .3450
                /note="central position to predicted promoter: -20;genetic
                evidence for the site"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to predicted promoter:50"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to bioA promoter:50"
                /bound_moiety="BioB documented site"
promoter        3413. .3441
                /note="factor Sigma70; promoter bioB; documented +1 at
                808525"
promoter        complement(3447. .3473)
                /note="factor Sigma70; promoter bioA; documented +1 at
                808515"
gene            3493. .4533
                /gene="bioB"
                /note="b0775"
CDS             3493. .4533
                /gene="bioB"
                /EC_number="2.8.1.-"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                /codon_start=1
                /transl_table=11
                /product="biotin synthesis, sulfur insertion?"
                /protein_id="AAC73862.1"
                /db_xref="GI:1786992"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            4530. .5684
                /gene="bioF"
                /note="b0776"
CDS             4530. .5684
                /gene="bioF"
                /EC_number="2.3.1.47"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                /codon_start=1
                /transl_table=11
                /product="8-amino-7-oxononanoate synthase"
                /protein_id="AAC73863.1"
                /db_xref="GI:1786993"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
```

```
  Query Match             100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity   100.0%;  Pred. No. 0.36;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||||||||||||||||||||||||
Db    3582 GCATCGCCAGCATTTCGATCCTCGT 3606


RESULT  13
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
```

```
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .13501
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .7576
                     /note="O-island #36; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655; Cryptic prophage
                     CP-933K; includes one copy of the 13 bp direct repeat that
                     flanks the prophage"
     gene            66. .665
                     /gene="lomK"
                     /note="Z0981"
     CDS             66. .665
                     /gene="lomK"
                     /function="putative membrane; Other or unknown (Phage or
                     Prophage Related)"
                     /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                     residues 1 to 199 of 199 from GenPept 118 :
                     gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
                     [prophage P-EibA]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein of prophage
                     CP-933K"
                     /protein_id="AAG55137.1"
                     /db_xref="GI:12513752"
                     /translation="MRKVCAAILSAATCLAVSGVPAWASEHQSTLSAGYLHASTDAPG
                     SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG
                     PSVRVNEWFSAYAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW
                     GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"
     gene            719. .2041
                     /gene="Z0982"
     CDS             719. .2041
                     /gene="Z0982"
                     /function="putative structure; Structural component (Phage
                     or Prophage Related)"
                     /note="Residues 164 to 440 of 440 are 68.79 pct identical
                     to residues 381 to 645 of 645 from GenPept 118 :
                     gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
                     tail fiber protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tail component of prophage CP-933K"
                     /protein_id="AAG55138.1"
                     /db_xref="GI:12513753"
                     /translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE
                     NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQPGTLNDFLGAMTEDDAR
                     PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE
                     SARQAAESAASAKKSEEASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA
                     TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE
                     RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA
                     GPQGPKGETGAAGPVGATGPQGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI
                     TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"
     gene            2004. .2312
                     /gene="Z0984"
     CDS             2004. .2312
                     /gene="Z0984"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 102 of 102 are 98.03 pct identical to
                     residues 28 to 129 of 129 from GenPept 118 :
                     gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
                     hypothetical protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55139.1"
                     /db_xref="GI:12513754"
                     /translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR
                     LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK
                     "
     gene            2489. .3469
                     /gene="Z0985"
     CDS             2489. .3469
                     /gene="Z0985"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 13 to 326 of 326 are 55.73 pct identical
                     to residues 18 to 331 of 336 from GenPept 118 :
                     gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
                     predicted by Glimmer [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55140.1"
                     /db_xref="GI:12513755"
                     /translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF
                     QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS
                     KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR
                     EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD
                     MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD
                     AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"
     gene            3530. .4522
                     /gene="Z0986"
     CDS             3530. .4522
                     /gene="Z0986"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55141.1"
                     /db_xref="GI:12513756"
                     /translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
                     SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK
                     VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY
                     VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG
                     WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT
                     EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"
     gene            5350. .6231
                     /gene="Z0989"
     CDS             5350. .6231
                     /gene="Z0989"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55142.1"
                     /db_xref="GI:12513757"
                     /translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV
```

```
                     KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGEISRKIEVMLNASPVSAAM
                     ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                     FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDLIGIRMDKINGESLLNISSLPAQ
                     AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                     SYHGGKQDLISVVLSKI"
     gene            6462. .7160
                     /gene="Z0990"
     CDS             6462. .7160
                     /gene="Z0990"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55143.1"
                     /db_xref="GI:12513758"
                     /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLRF
                     QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                     IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                     SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                     VSLGFQQVRLHPLL"
     gene            complement(7668. .8144)
                     /gene="ybhB"
                     /note="Z0992"
     CDS             complement(7668. .8144)
                     /gene="ybhB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                     residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                     MG1655: B0773"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55144.1"
                     /db_xref="GI:12513759"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(8203. .9492)
                     /gene="bioA"
                     /note="Z0993"
     CDS             complement(8203. .9492)

  Query Match             100.0%;  Score 25;  DB 1;  Length 13501;
  Best Local Similarity   100.0%;  Pred. No. 0.36;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcgt 25
          |||||||||||||||||||||||||
Db   9668 GCATCGCCAGCATTTCGATCCTCGT 9692


RESULT  14
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .1245
                     /gene="ECs0753"
     CDS             79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
     gene            1245. .2114
                     /gene="ECs0754"
     CDS             1245. .2114
                     /gene="ECs0754"
                     /note="similar to SUCD_ECOLI gi|1786949 percent identity
                     100 in 289 aa (Conserved in E.coli K-12)"
                     /codon_start=1
```

```
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase alpha subunit"
                     /protein_id="BAB34177.1"
                     /db_xref="GI:13360213"
                     /translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
                     GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
                     LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
                     AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
                     AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
                     LADIGEALKTVLK"
gene                 2781. .3686
                     /gene="ECs0755"
CDS                  2781. .3686
                     /gene="ECs0755"
                     /note="probable transcriptional regulator,similar to
                     transcriptional regulators e.g. glycine cleavage system
                     transcription activator (gcv operon activator) -
                     Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
                     identity 31 in 300 aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative transcription regulator"
                     /protein_id="BAB34178.1"
                     /db_xref="GI:13360214"
                     /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
                     EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
                     PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
                     LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
                     NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
                     EKSDSPHLQQFREWLLAKSVVPQEM"
gene                 complement(3720. .4322)
                     /gene="ECs0756"
CDS                  complement(3720. .4322)
                     /gene="ECs0756"
                     /note="probable cob(I)alamin adenosyltransferase,similar
                     to cob(I)alamin adenosyltransferases (corrinoid
                     adenosyltransferases) e.g. [Escherichia coli]
                     gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
                     aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative cob(I)alamin adenosyltransferase"
                     /protein_id="BAB34179.1"
                     /db_xref="GI:13360215"
                     /translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
                     KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
                     QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
                     IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"
gene                 complement(4332. .5984)
                     /gene="ECs0757"
CDS                  complement(4332. .5984)
                     /gene="ECs0757"
                     /note="probable fumarate hydratase,similar to fumarate
                     hydratases e.g. fumarate hydratase class I, aerobic
                     (fumarase) - Escherichia coli
                     gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
                     aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative fumarate hydratase"
                     /protein_id="BAB34180.1"
                     /db_xref="GI:13360216"
                     /translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
                     ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
                     NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
                     TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
                     GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
                     EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
                     EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
                     DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
                     QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSIGGAAALLAQEYVKSLRCLEYPELG
                     MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"
gene                 complement(6092. .7372)
                     /gene="ECs0758"
CDS                  complement(6092. .7372)
                     /gene="ECs0758"
                     /note="probable transport protein,similar to
                     glutamate/aspartate transport proteins (proton glutamate
                     symport proteins) e.g. [Bacillus stearothermophilus]
                     gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
                     aa,also similar to C4-dicarboxylate transpor"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative transport protein"
                     /protein_id="BAB34181.1"
                     /db_xref="GI:13360217"
                     /translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
                     TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
                     NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
                     FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
                     VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
                     GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
                     LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
                     AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"
gene                 complement(7533. .7850)
                     /gene="ECs0759"
CDS                  complement(7533. .7850)
                     /gene="ECs0759"
                     /note="unknown"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="hypothetical protein"
                     /protein_id="BAB34182.1"

  Query Match             100.0%;  Score 25;  DB 1;  Length 297816;
  Best Local Similarity   100.0%;  Pred. No. 0.35;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||||||||||||||||||||||||
Db   92400 GCATCGCCAGCATTTCGATCCTCGT 92424


RESULT  15
AF250776
LOCUS       AF250776     5526 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
            modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
            biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
            biotin biosynthesis protein BioC (bioC) genes, complete cds; and
            dethiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION   AF250776
VERSION     AF250776.1  GI:12620124
KEYWORDS    .
SOURCE      uncultured bacterium pCosHE2.
  ORGANISM  uncultured bacterium pCosHE2
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
```

```
     source          1. .5526
                     /organism="uncultured bacterium pCosHE2"
                     /db_xref="taxon:143797"
                     /clone="pCosHE2"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     CDS             complement(52. .528)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical 17.1 kDa protein in modC-bioA
                     intergenic region"
                     /protein_id="AAG60577.1"
                     /db_xref="GI:12620125"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKAG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(587. .1876)
                     /gene="bioA"
     CDS             complement(587. .1876)
                     /gene="bioA"
                     /note="7,8-diaminopelargonic acid
                     synthetase-aminotransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="DAPA-aminotransferase BioA"
                     /protein_id="AAG60578.1"
                     /db_xref="GI:12620126"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LKCVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            1963. .3003
                     /gene="bioB"
     CDS             1963. .3003
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase BioB"
                     /protein_id="AAG60579.1"
                     /db_xref="GI:12620127"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPNERDMPYLEQMVQGVKDLGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3000. .4154
                     /gene="bioF"
     CDS             3000. .4154
                     /gene="bioF"
                     /note="8-amino-7-oxononanoate synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="KAPA synthetase BioF"
                     /protein_id="AAG60580.1"
                     /db_xref="GI:12620128"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQLVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAAL
                     RASLAVIRRDEGDARREKLVSLIAHFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4141. .4896
                     /gene="bioC"
     CDS             4141. .4896
                     /gene="bioC"
                     /note="reaction step prior to synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin biosynthesis protein BioC"
                     /protein_id="AAG60581.1"
                     /db_xref="GI:12620129"
                     /translation="MATVNKQAIAAAFGRAAALYEQHADLQRQSADVLLAMLPERKYT
                     HVLDAGCGPGWMTRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRSKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4883. .>5526
                     /gene="bioD"
     CDS             4883. .>5526
                     /gene="bioD"
                     /note="DTB synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase BioD"
                     /protein_id="AAG60582.1"
                     /db_xref="GI:12620130"
                     /translation="MLVSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVAS
                     GSEKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESSVMS
                     SGLRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAML
                     TAQAIQHAGLTLAGWVANDVTPPGKRHAEYITTLTRMIPAPLLGEIPWLAENPE"
BASE COUNT      1274 a    1507 c    1567 g    1178 t
ORIGIN


  Query Match           96.0%;  Score 24;  DB 1;  Length 5526;
  Best Local Similarity  100.0%;  Pred. No. 1.1;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcgt 25
           ||||||||||||||||||||||||
Db    2053 CATCGCCAGCATTTCGATCCTCGT 2076


RESULT  16
AE003860
LOCUS       AE003860    10593 bp    DNA             BCT       15-JUN-2001
DEFINITION  Xylella fastidiosa 9a5c, section 6 of 229 of the complete genome.
ACCESSION   AE003860 AE003849
VERSION     AE003860.1  GI:9104830
KEYWORDS    .
SOURCE      Xylella fastidiosa 9a5c.
  ORGANISM  Xylella fastidiosa 9a5c
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xylella.
REFERENCE   1  (bases 1 to 10593)
  AUTHORS   Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
            Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrer,H.,
            Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
            Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H.,
            Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
            Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
            Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
            Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
            and Marino,C.L.
  TITLE     The genome sequence of the plant pathogen Xylella fastidiosa. The
            Xylella fastidiosa Consortium of the Organization for Nucleotide
            Sequencing and Analysis
  JOURNAL   Nature 406 (6792), 151-157 (2000)
  MEDLINE   20365717
  PUBMED    10910347
REFERENCE   2  (bases 1 to 10593)
  AUTHORS   Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
            Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
            Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
```

```
            Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
            Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,
            Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
            Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
            Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
            Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
            Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
            Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
            Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
            Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
            Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
            Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
            Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
            Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
            Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
            de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
            Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
            Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
            Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
            Tsuhako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
            Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
            Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
            13083-970, Brazil
FEATURES            Location/Qualifiers
     source         1. .10593
                    /organism="Xylella fastidiosa 9a5c"
                    /db_xref="taxon:160492"
                    /clone="9a5c"
     gene           complement(277. .1356)
                    /gene="XF0061"
     CDS            complement(277. .1356)
                    /gene="XF0061"
                    /note="similar to SP|P07674 (percent identity: 28 %/query
                    alignment coverage:  64.9 %/subject alignment coverage:
                    65.1 %); identified by sequence similarity; putative; ORF
                    located using Glimmer/RBSfinder"
                    /codon_start=1
                    /transl_table=11
                    /product="transcriptional repressor"
                    /protein_id="AAF82874.1"
                    /db_xref="GI:9104831"
                    /translation="MLLLQTLFELDRATAMSLFGNRIKPTQEQNQLATVAFSGEHIEE
                    IDPELIDLEKQMRSKDNPGFTIESLMELGNDMKRDGQHEPVILRKNSKKPGRYLMVAG
                    ERRWLACKAVDIKLKAVIRELTDDQVRRVQRAENIQRENLTQLEIAVALRDDKERLGT
                    LEKVAAEWNKGINWVSERINFLEVVETNSLTSQAVSEGITADITAINDLYRLEKVNQT
                    AAKTIIDQAKVDPGINMRKTVREKLQNAKPTTNEKNATTDSRMQESKTDFQEGETHLQ
                    NISNQEKPATVASAPLPPDAAQLVALAAKQMETLCTEVSALSKILHQRQDVISALGDV
                    LQNVVQAQAALFRTRHVLASLANSK"
     gene           complement(1331. .1657)
                    /gene="XF0062"
     CDS            complement(1331. .1657)
                    /gene="XF0062"
                    /note="hypothetical protein; identified by sequence
                    similarity; putative; ORF located using Glimmer/RBSfinder"
                    /codon_start=1
                    /transl_table=11
                    /product="hypothetical protein"
                    /protein_id="AAF82875.1"
                    /db_xref="GI:9104832"
                    /translation="MYHRSIFQRSNSIHNETVKHKCWSDLQNHDIDTTYPLPNSTNNI
                    LEWDQNFKNTNEICAENTSPIQSIYRPRPPEQHVNIFPILIPHLMPVQHLKMYQARHA
                    VVTDIV"
     gene           complement(2427. .3152)
                    /gene="XF0063"
     CDS            complement(2427. .3152)
                    /gene="XF0063"
                    /note="similar to GI|5921548 (percent identity: 37 %/query
                    alignment coverage:  93.4 %/subject alignment coverage:
                    95.7 %); identified by sequence similarity; putative; ORF
                    located using Glimmer/RBSfinder"
                    /codon_start=1
                    /transl_table=11
                    /product="competence protein F"
                    /protein_id="AAF82876.1"
                    /db_xref="GI:9104833"
                    /translation="MKDAVNSNHILPDTWCDRLLRWLLPSYCLACNEPGHQGKTLCAT
                    CQLLLPRMENACLRCAAPLKELEGIHPTCGPCLYDPPPLETAHAAFLYHWPINKLISR
                    FKFHQDLACGHLLADNMAMAAQHWPRPQALVPVSLHHGRLRTRGYDQALELARHVGNR
                    LQLPCLPALRRVRATLPQSQLNANKRRSNLHNAFIPIHPLPAHVALVDDVMTTGTTLH
                    AAARALRHTGVTRIDAWICARVP"
     gene           3202. .4227
                    /gene="XF0064"
     CDS            3202. .4227
                    /gene="XF0064"
                    /note="similar to SP|P36569 (percent identity: 60 %/query
                    alignment coverage:  94.4 %/subject alignment coverage:
                    93.1 %); identified by sequence similarity; putative; ORF
                    located using Glimmer/RBSfinder/Start codon shift: 66"
                    /codon_start=1
                    /transl_table=11
                    /product="biotin synthase"
                    /protein_id="AAF82877.1"
                    /db_xref="GI:9104834"
                    /translation="MSSIIRYDWTAEELHALFDLSLPELLYRAASVHRQHFDPAEIQV
                    STLLSVKTGGCPEDCSYCPQAQRYDTGVTAQKLMDVDAVVAKARQAKLAGASRFCMGA
                    AWRSPKDRDIPKIAAMIREVKALGLETCATLGMLNTCQAQALKDAGLDYYNHNVDTSP
                    DFYDSVIHTRQYQDRLDTLAHVRDVGLKTCCGGIVGMGETRQHRVGLLLTLATLPAHP
                    DSVPVNLLVQVAGTPLHGTQTLDPFEFVRMIAVARITMPRSMVRLSAGRESMSDELQL
                    LCFMAGANSIFYGEKLLTTANPETERDQALFQRLGLRPMHLMENVSNQDQHHGNVHAD
                    IACKHVV"
     gene           4281. .4382
                    /gene="XF0065"
     CDS            4281. .4382
                    /gene="XF0065"
                    /note="hypothetical protein; identified by sequence
                    similarity; putative; ORF located using Glimmer/RBSfinder"
                    /codon_start=1
                    /transl_table=11
                    /product="hypothetical protein"
                    /protein_id="AAF82878.1"
                    /db_xref="GI:9104835"
                    /translation="MATLVDVVSFMYDGVNLDSWSDGGERRKVVEFS"
     gene           4486. .5673
                    /gene="XF0066"
     CDS            4486. .5673
                    /gene="XF0066"
                    /note="similar to SP|O34731 (percent identity: 33 %/query
                    alignment coverage:  63.3 %/subject alignment coverage:
                    96.2 %); identified by sequence similarity; putative; ORF
                    located using Glimmer/RBSfinder"
                    /codon_start=1
                    /transl_table=11
                    /product="conserved hypothetical protein"
                    /protein_id="AAF82879.1"
                    /db_xref="GI:9104836"
                    /translation="MSLINHKNGVSCNCGSTCSRDILNRLLWRRVVSGCNRYPVHLSY
                    LALSSMCFTFRCSFLATIVVALLTACSSAPRSASRLSSVSQLSDISLPPLRIGIALGG
                    GAAKGFAHIGVLKMLEANGLAPSVVAGTSAGSVVGALYASGMDSFKIQTKAVAMDESN
                    IRDLRFFSGGLVQGQKLQNYVNELVGKRPIEKLNKPFGAIATRLEDGQRTMFVRGNVG
                    EAVRASCSIPGVFEPVTIGAYHYVDGGITSPVPVDAVRQLGADFVIAVDISSKATGKN
                    PGGLVGTVNQSISIMGQHLGEAELKRANIVIHPKVLDIGSTDFDQRNAAILEGERATI
                    AVIPEIREKLAQLQQQRIQAARKTANQSAIEAVESYRRCQNSRTRWEKLRGKEKDCTK
                    SSL"
     gene           6138. .6377
                    /gene="XF0067"
     CDS            6138. .6377
                    /gene="XF0067"
                    /note="hypothetical protein; identified by sequence
                    similarity; putative; ORF located using Glimmer/RBSfinder"
                    /codon_start=1
                    /transl_table=11
                    /product="hypothetical protein"
                    /protein_id="AAF82880.1"
                    /db_xref="GI:9104837"
```

```
                     /translation="MSCKEIFEYIVIKQCWLPMFFAQMGMVMGAIWCADRYPPIVLIS
                     RCVLSYYVGWGYRVRFSVILWPMCYWLGVMIVLAF"
     gene            complement(6494. .7495)
                     /gene="XF0068"
     CDS             complement(6494. .7495)
                     /gene="XF0068"
                     /note="similar to SP|P26601 (percent identity: 52 %/query
                     alignment coverage:  86.8 %/subject alignment coverage:
                     99.7 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hydroxybenzoate octaprenyltransferase"
                     /protein_id="AAF82881.1"
                     /db_xref="GI:9104838"
                     /translation="MILNQEHEHIQVYNHHLNSFQFPLNPSSDNSPVSMAYERFTSAI
                     TLLMHWRNRLDPYWKLARGDRPVGFLLLLWPTWWALWLAAGGVPPWWTLCVFTTGIWL
                     TRSAGCVINDYTDRWLDPHVERTRIRPLVTGAVSPRNALLMFATLMLIAFGLVLTMNQ
                     LTVLLSVVGLFLAMTYPYLKRYTYLPQVYLGLAFGWGIPMAFAAIQGKVPTQAWLLYV
                     ANILWTTAYDTWCAMVDRDDDIKMGAKSTAILFADLDLTVQGVLYTLMLFTLCLVGLR
                     ATLSHTYWISLIAAVALIGYQFIIARRREPTACFRAFMHNNWVGMTIFAGIALATTH"

  Query Match             81.6%;  Score 20.4;  DB 1;  Length 10593;
  Best Local Similarity   95.5%;  Pred. No. 52;
  Matches   21;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcct 22
           |||||| ||||||||||||||||
Db    3297 GCATCGTCAGCATTTCGATCCT 3318


RESULT  17
AF250770
LOCUS       AF250770      965 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosHE1 DAPA-aminotransferase (bioA) and
            biotin synthase (bioB) genes, partial cds.
ACCESSION   AF250770
VERSION     AF250770.1  GI:12620104
KEYWORDS    .
SOURCE      uncultured bacterium pCosHE1.
  ORGANISM  uncultured bacterium pCosHE1
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 965)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 965)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .965
                     /organism="uncultured bacterium pCosHE1"
                     /db_xref="taxon:143796"
                     /clone="pCosHE1"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     gene            complement(<1. .438)
                     /gene="bioA"
     CDS             complement(<1. .438)
                     /gene="bioA"
                     /note="7,8-diaminopelargonic acid
                     synthetase-aminotransferase; similar to BioA"
                     /codon_start=1
                     /transl_table=11
                     /product="DAPA-aminotransferase"
                     /protein_id="AAG60563.1"
                     /db_xref="GI:12620105"
                     /translation="MTQDDLAFDQQHIWHPYTSMTRPXPVYPVASAXAHCQRRRRLVD
                     GMSSWWAAIHGYNHPRLNAAMKAQIEQMSHVMFDGITHQPAVDLCRRXVAMTPDALEC
                     VFLADSGSVAXEVXDENGAASTAPLPADRPPERYHXAFFIPXLQ"
     gene            525. .>965
                     /gene="bioB"
     CDS             525. .>965
                     /gene="bioB"
                     /note="similar to BioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase"
                     /protein_id="AAG60564.1"
                     /db_xref="GI:12620106"
                     /translation="MXHHARWTMSQVTELFNKPFLELMFEAQQVHRQHFDPRHVQXST
                     LLSIKTGACPEDCKYCPQSARYKTGLESERLMEVEQVLDSARKAKNAGSTRFCMGAAW
                     KKPHTIVTCPYLEQMVKGVKEMGLEALYDPRHAERRAXXSACLLA"
BASE COUNT      230 a    256 c    285 g    180 t     14 others
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 1;  Length 965;
  Best Local Similarity   87.5%;  Pred. No. 2e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcg 24
           |||||| ||||| |||||||| ||
Db     614 GCATCGTCAGCACTTCGATCCGCG 637


RESULT  18
AF250768
LOCUS       AF250768      4000 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosFS1 BioA-like protein (bioA) gene, partial
            cds; BioB-like protein (bioB) and KAPA-synthetase (bioF) genes,
            complete cds; and biotin biosynthesis protein BioC (bioC) gene,
            partial cds.
ACCESSION   AF250768
VERSION     AF250768.1  GI:12620095
KEYWORDS    .
SOURCE      uncultured bacterium pCosFS1.
  ORGANISM  uncultured bacterium pCosFS1
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 4000)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 4000)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .4000
                     /organism="uncultured bacterium pCosFS1"
                     /db_xref="taxon:143795"
                     /clone="pCosFS1"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     gene            complement(<1. .1177)
                     /gene="bioA"
     CDS             complement(<1. .1177)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="BioA-like protein"
                     /protein_id="AAG60558.1"
                     /db_xref="GI:12620098"
                     /translation="MSVTDSDLEFDQRHIWHPYTSMSHPLPCYPVEAASGVELQLADG
                     RRLIDGMSSWWAAIHGYNHPQLNLAATQQLEKMSHVMFGGITHPAANELCRKVVAMTP
                     EALQCVFLADSGSVAVEVSLKMALQYWQARGERRQRILTLRHGYHGDTFGAMSVCDPD
                     NSMHSLYQGYLAPHLFATAPQCGFDQAWDENDIAPFAALLEQHAGEVAAVILEPVVQG
```

```
                AGGMRIYHPTYLERVRELCDRYQVLLIADEIATGFGRTGKLFACEHAEVVPDILCLGK
                ALTGGYMTLSATLTTRHVAETISNGAAGCFMHGPTFMGNPLACAVANASLSLLAENRW
                QQQVAAIETQLKQQLLPLKMQPKVADVRVLGAIGVVEMREPVDMAQLQRGFVERGVWI
                "
     gene       1265. .2305
                /gene="bioB"
     CDS        1265. .2305
                /gene="bioB"
                /codon_start=1
                /transl_table=11
                /product="BioB-like protein"
                /protein_id="AAG60559.1"
                /db_xref="GI:12620099"
                /translation="MMADRIHWTVGLAQTLFDKPLLELLFEAQTVHRQHFDPRQVQVS
                TLLSIKTGACPEDCKYCPQSSRYKTGLESERLMQVEQVLESARKAKANGSTRFCMGAA
                WKNPHERDMPYLQQMVQGVKAMGMETCMTLGTLDGSQAERLAEAGLDYYNHNLDTSPE
                FYGNIITTRSYQERLDTLGKVRGAGIKVCSGGIVGLGETVXXRAGLLVQLANLPTPPE
                SVPINMLVKVKGTPLADNDDVDPFDFIRTIAVARIMMPSSYVRLSAGREQMNEQTQAM
                CFMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTETEHGDNQQQEVLAARLM
                QADTAEFYNAAL"
     gene       2305. .3456
                /gene="bioF"
     CDS        2305. .3456
                /gene="bioF"
                /note="7-keto-8-aminopelargonic acid synthetase; similar
                to BioF"
                /codon_start=1
                /transl_table=11
                /product="KAPA-synthetase"
                /protein_id="AAG60556.1"
                /db_xref="GI:12620096"
                /translation="MSWQQRIEQALEERXQSAAYRHRLVIKGGNGRYLQLNDQRYLNF
                SGNDYLGLSQNAQVIAAWQQGAQQYGVGSGGSGHVTGFCTAHQALEQQLAQWLGYPRS
                LLFISGYAANQALLAALMQGEDRILADRLSHASLLEAAAHSPAQLRRFRHNQPQALAD
                LLAKPCSGQTLAVTEGVFSMDGDSAPLAELYRLTRDAGAWLMVDDAHGIGVQGEQGRG
                SCWQQGVRPELLVVTFGKAFGLSGAAVLCDESTAEYLLQFARHLIYSTAMPPAQACAL
                QAALHCIQQGDELRTRLNTNIQRFRQGAAQLSLTLTESLTAIQPLLVGDNQRALDLAQ
                HLRGQGLWVSAIRPPTVPPGGARLRITLTAAHQPQDIDRLLEVLHDVSV"
     gene       3440. .>4000
                /gene="bioC"
     CDS        3440. .>4000
                /gene="bioC"
                /note="similar to BioC; probably reaction step prior to
                synthesis of pimeloyl-CoA"
                /codon_start=1
                /transl_table=11
                /product="biotin biosynthesis protein BioC"
                /protein_id="AAG60557.1"
                /db_xref="GI:12620097"
                /translation="MTSAFEQVNKQAVASAFSRAAVGYDAAAVLQRDVGERLLGMGQE
                HQGEQLLDAGCGTGYFSRRWRELGKQVTALDLAPGMLAFARQQQAADHYLLGDIENVP
                LGDAAVDISFSSLVVQWCSDLPRALAELYRVTRPGGVILFSTLAEGSLHELGDAWQQV
                DGERHVNDFLPLAQIEAACAGYRHHLQ"
BASE COUNT      794 a   1164 c   1236 g    802 t      4 others
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 1;  Length 4000;
  Best Local Similarity   87.5%;  Pred. No. 1.9e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcgt 25
           || |||||||||||| ||||| |||
Db    1358 CACCGCCAGCATTTTGATCCGCGT 1381


RESULT 19
AC016014/c
LOCUS       AC016014    68419 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens clone RP11-26J10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC016014
VERSION     AC016014.2  GI:9134660
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 68419)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-26J10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68419)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6449522.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4643
                Center clone name: 26_J_10
            --------------.
            * NOTE: This record contains 76 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      788: contig of 788 bp in length
            *      789 888: gap of      100 bp
            *      889     1666: contig of 778 bp in length
            *     1667 1766: gap of      100 bp
            *     1767     2567: contig of 801 bp in length
            *     2568 2667: gap of      100 bp
            *     2668     3636: contig of 969 bp in length
            *     3637 3736: gap of      100 bp
            *     3737     4595: contig of 859 bp in length
            *     4596 4695: gap of      100 bp
            *     4696     5517: contig of 822 bp in length
            *     5518 5617: gap of      100 bp
            *     5618     6436: contig of 819 bp in length
            *     6437 6536: gap of      100 bp
            *     6537     7361: contig of 825 bp in length
            *     7362 7461: gap of      100 bp
            *     7462     8283: contig of 822 bp in length
            *     8284 8383: gap of      100 bp
            *     8384     9171: contig of 788 bp in length
            *     9172 9271: gap of      100 bp
            *     9272    10074: contig of 803 bp in length
            *    10075 10174: gap of      100 bp
            *    10175    10962: contig of 788 bp in length
```

```
*    10963 11062: gap of      100 bp
*    11063    11845: contig of 783 bp in length
*    11846 11945: gap of      100 bp
*    11946    12758: contig of 813 bp in length
*    12759 12858: gap of      100 bp
*    12859    13619: contig of 761 bp in length
*    13620 13719: gap of      100 bp
*    13720    14542: contig of 823 bp in length
*    14543 14642: gap of      100 bp
*    14643    15428: contig of 786 bp in length
*    15429 15528: gap of      100 bp
*    15529    16328: contig of 800 bp in length
*    16329 16428: gap of      100 bp
*    16429    17221: contig of 793 bp in length
*    17222 17321: gap of      100 bp
*    17322    18099: contig of 778 bp in length
*    18100 18199: gap of      100 bp
*    18200    18980: contig of 781 bp in length
*    18981 19080: gap of      100 bp
*    19081    19869: contig of 789 bp in length
*    19870 19969: gap of      100 bp
*    19970    20807: contig of 838 bp in length
*    20808 20907: gap of      100 bp
*    20908    21769: contig of 862 bp in length
*    21770 21869: gap of      100 bp
*    21870    22649: contig of 780 bp in length
*    22650 22749: gap of      100 bp
*    22750    23545: contig of 796 bp in length
*    23546 23645: gap of      100 bp
*    23646    24442: contig of 797 bp in length
*    24443 24542: gap of      100 bp
*    24543    25345: contig of 803 bp in length
*    25346 25445: gap of      100 bp
*    25446    26227: contig of 782 bp in length
*    26228 26327: gap of      100 bp
*    26328    27126: contig of 799 bp in length
*    27127 27226: gap of      100 bp
*    27227    28039: contig of 813 bp in length
*    28040 28139: gap of      100 bp
*    28140    28965: contig of 826 bp in length
*    28966 29065: gap of      100 bp
*    29066    29854: contig of 789 bp in length
*    29855 29954: gap of      100 bp
*    29955    30736: contig of 782 bp in length
*    30737 30836: gap of      100 bp
*    30837    31593: contig of 757 bp in length
*    31594 31693: gap of      100 bp
*    31694    32483: contig of 790 bp in length
*    32484 32583: gap of      100 bp
*    32584    33391: contig of 808 bp in length
*    33392 33491: gap of      100 bp
*    33492    34285: contig of 794 bp in length
*    34286 34385: gap of      100 bp
*    34386    35223: contig of 838 bp in length
*    35224 35323: gap of      100 bp
*    35324    36128: contig of 805 bp in length
*    36129 36228: gap of      100 bp
*    36229    37039: contig of 811 bp in length
*    37040 37139: gap of      100 bp
*    37140    37899: contig of 760 bp in length
*    37900 37999: gap of      100 bp
*    38000    38779: contig of 780 bp in length
*    38780 38879: gap of      100 bp
*    38880    39673: contig of 794 bp in length
*    39674 39773: gap of      100 bp
*    39774    40594: contig of 821 bp in length
*    40595 40694: gap of      100 bp
*    40695    41475: contig of 781 bp in length
*    41476 41575: gap of      100 bp
*    41576    42363: contig of 788 bp in length
*    42364 42463: gap of      100 bp
*    42464    43252: contig of 789 bp in length
*    43253 43352: gap of      100 bp
*    43353    44179: contig of 827 bp in length
*    44180 44279: gap of      100 bp
*    44280    45104: contig of 825 bp in length
*    45105 45204: gap of      100 bp
*    45205    46034: contig of 830 bp in length
*    46035 46134: gap of      100 bp
*    46135    46936: contig of 802 bp in length
*    46937 47036: gap of      100 bp
*    47037    47869: contig of 833 bp in length
*    47870 47969: gap of      100 bp
*    47970    48770: contig of 801 bp in length
*    48771 48870: gap of      100 bp
*    48871    49668: contig of 798 bp in length
*    49669 49768: gap of      100 bp
*    49769    50560: contig of 792 bp in length
*    50561 50660: gap of      100 bp
*    50661    51418: contig of 758 bp in length
*    51419 51518: gap of      100 bp
*    51519    52301: contig of 783 bp in length
*    52302 52401: gap of      100 bp
*    52402    53189: contig of 788 bp in length
*    53190 53289: gap of      100 bp
*    53290    54081: contig of 792 bp in length
*    54082 54181: gap of      100 bp
*    54182    54984: contig of 803 bp in length
*    54985 55084: gap of      100 bp
*    55085    55863: contig of 779 bp in length
*    55864 55963: gap of      100 bp
*    55964    56778: contig of 815 bp in length
*    56779 56878: gap of      100 bp
*    56879    57657: contig of 779 bp in length
*    57658 57757: gap of      100 bp
*    57758    58554: contig of 797 bp in length
*    58555 58654: gap of      100 bp
*    58655    59472: contig of 818 bp in length
*    59473 59572: gap of      100 bp
*    59573    60360: contig of 788 bp in length
*    60361 60460: gap of      100 bp
*    60461    61240: contig of 780 bp in length
*    61241 61340: gap of      100 bp
*    61341    62195: contig of 855 bp in length
*    62196 62295: gap of      100 bp
*    62296    63086: contig of 791 bp in length
*    63087 63186: gap of      100 bp
*    63187    63978: contig of 792 bp in length
*    63979 64078: gap of      100 bp
*    64079    64874: contig of 796 bp in length
*    64875 64974: gap of      100 bp
*    64975    65754: contig of 780 bp in length

  Query Match              76.8%;  Score 19.2;  DB 2;  Length 68419;
  Best Local Similarity    87.5%;  Pred. No. 1.8e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcg 24
           ||||| ||| ||||||||||| ||
Db   28757 GCATCCCCACCATTTCGATCCACG 28734


RESULT  20
AC084172
LOCUS       AC084172   116334 bp    DNA             HTG       09-MAY-2001
DEFINITION  Homo sapiens chromosome 8 clone CTD-2577H2 map 8, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC084172
VERSION     AC084172.2  GI:12658056
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 116334)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
```

```
  TITLE     Homo sapiens chromosome 8, clone CTD-2577H2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 116334)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 3, 2001 this sequence version replaced gi:10801400.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11377
                Center clone name: 2577_H_2
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 113537 bases at least Q40
                Consensus quality: 114958 bases at least Q30
                Consensus quality: 115443 bases at least Q20
                Insert size: 115000; agarose-fp
                Insert size: 115734; sum-of-contigs
                Quality coverage: 7.3 in Q20 bases; agarose-fp
                Quality coverage: 7.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1008: contig of 1008 bp in length
            *    1009 1108: gap of      100 bp
            *    1109     5806: contig of 4698 bp in length
            *    5807 5906: gap of      100 bp
            *    5907    11021: contig of 5115 bp in length
            *   11022 11121: gap of      100 bp
            *   11122    16878: contig of 5757 bp in length
            *   16879 16978: gap of      100 bp
            *   16979    24688: contig of 7710 bp in length
            *   24689 24788: gap of      100 bp
            *   24789    56381: contig of 31593 bp in length
            *   56382 56481: gap of      100 bp
            *   56482   116334: contig of 59853 bp in length.
FEATURES             Location/Qualifiers
     source          1. .116334
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="CTD-2577H2"
                     /clone_lib="CTTD Human BAC"
     misc_feature    1. .1008
                     /note="assembly_fragment"
     misc_feature    1109. .5806
                     /note="assembly_fragment"
     misc_feature    5907. .11021
                     /note="assembly_fragment"
     misc_feature    11122. .16878
                     /note="assembly_fragment"
     misc_feature    16979. .24688
                     /note="assembly_fragment"
     misc_feature    24789. .56381
                     /note="assembly_fragment"
     misc_feature    56482. .116334
                     /note="assembly_fragment"
BASE COUNT    33562 a  23964 c  24393 g  33811 t    604 others
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 2;  Length 116334;
  Best Local Similarity   87.5%;  Pred. No. 1.8e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcgt 25
           |||| |||||||| |||||||||  |
Db 102265 CATCACCAGCATCTCGATCCTCCT 102288


RESULT  21
AC026294/c
LOCUS       AC026294    186616 bp    DNA              HTG       01-SEP-2000
DEFINITION  Homo sapiens chromosome 8 clone RP11-705C3, WORKING DRAFT SEQUENCE,
            15 unordered pieces.
ACCESSION   AC026294
VERSION     AC026294.3  GI:9958420
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186616)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186616)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:8099313.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0705C03
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 177884 bases at least Q40
            Consensus quality: 180695 bases at least Q30
            Consensus quality: 182282 bases at least Q20
            Insert size: 199000; agarose-fp
```

```
            Insert size: 185216; sum-of-contigs
            Quality coverage: 4.02 in Q20 bases; agarose-fp
            Quality coverage: 4.35 in Q20 bases; sum-of-contigs
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2457: contig of 2457 bp in length
            *     2458     2557: gap of unknown length
            *     2558     6164: contig of 3607 bp in length
            *     6165     6264: gap of unknown length
            *     6265     8851: contig of 2587 bp in length
            *     8852     8951: gap of unknown length
            *     8952    12910: contig of 3959 bp in length
            *    12911    13010: gap of unknown length
            *    13011    17550: contig of 4540 bp in length
            *    17551    17650: gap of unknown length
            *    17651    23522: contig of 5872 bp in length
            *    23523    23622: gap of unknown length
            *    23623    28616: contig of 4994 bp in length
            *    28617    28716: gap of unknown length
            *    28717    39183: contig of 10467 bp in length
            *    39184    39283: gap of unknown length
            *    39284    47100: contig of 7817 bp in length
            *    47101    47200: gap of unknown length
            *    47201    59667: contig of 12467 bp in length
            *    59668    59767: gap of unknown length
            *    59768    73204: contig of 13437 bp in length
            *    73205    73304: gap of unknown length
            *    73305    87755: contig of 14451 bp in length
            *    87756    87855: gap of unknown length
            *    87856   105521: contig of 17666 bp in length
            *   105522   105621: gap of unknown length
            *   105622   138110: contig of 32489 bp in length
            *   138111   138210: gap of unknown length
            *   138211   186616: contig of 48406 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186616
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="RP11-705C3"
     misc_feature    1. .2457
                     /note="assembly_name:Contig30"
     misc_feature    2558. .6164
                     /note="assembly_name:Contig31"
     misc_feature    6265. .8851
                     /note="assembly_name:Contig32"
     misc_feature    8952. .12910
                     /note="assembly_name:Contig33"
     misc_feature    13011. .17550
                     /note="assembly_name:Contig34"
     misc_feature    17651. .23522
                     /note="assembly_name:Contig35"
     misc_feature    23623. .28616
                     /note="assembly_name:Contig36
                     clone_end:SP6
                     vector_side:right"
     misc_feature    28717. .39183
                     /note="assembly_name:Contig38
                     clone_end:T7
                     vector_side:right"
     misc_feature    39284. .47100
                     /note="assembly_name:Contig37"
     misc_feature    47201. .59667
                     /note="assembly_name:Contig39"
     misc_feature    59768. .73204
                     /note="assembly_name:Contig40"
     misc_feature    73305. .87755
                     /note="assembly_name:Contig41"
     misc_feature    87856. .105521
                     /note="assembly_name:Contig42"
     misc_feature    105622. .138110
                     /note="assembly_name:Contig43"
     misc_feature    138211. .186616
                     /note="assembly_name:Contig44"
BASE COUNT    53026 a  36754 c  37995 g  57255 t   1586 others
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 2;  Length 186616;
  Best Local Similarity   87.5%;  Pred. No. 1.8e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcgt 25
           |||| ||||||| |||||||||  |
Db   43327 CATCACCAGCATCTCGATCCTCCT 43304


RESULT  22
AC090154
LOCUS       AC090154   189025 bp    DNA              HTG       20-MAY-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-705C3 map 8, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC090154
VERSION     AC090154.2  GI:14150954
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 189025)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-705C3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189025)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 20, 2001 this sequence version replaced gi:12957783.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11757
```

```
           Center clone name: 705_C_3
          ---------- Summary Statistics
           Sequencing vector: Plasmid; n/a; 100% of reads
           Chemistry: Dye-terminator Big Dye; 100% of reads
           Assembly program: Phrap; version 0.960731
           Consensus quality: 185046 bases at least Q40
           Consensus quality: 186416 bases at least Q30
           Consensus quality: 187380 bases at least Q20
           Insert size: 190000; agarose-fp
           Insert size: 188125; sum-of-contigs
           Quality coverage: 9.0 in Q20 bases; agarose-fp
           Quality coverage: 9.1 in Q20 bases; sum-of-contigs
          ----------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 10 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1      978: contig of 978 bp in length
          *      979 1078: gap of      100 bp
          *     1079     1991: contig of 913 bp in length
          *     1992 2091: gap of      100 bp
          *     2092     4062: contig of 1971 bp in length
          *     4063 4162: gap of      100 bp
          *     4163    77746: contig of 73584 bp in length
          *    77747 77846: gap of      100 bp
          *    77847    83726: contig of 5880 bp in length
          *    83727 83826: gap of      100 bp
          *    83827    96215: contig of 12389 bp in length
          *    96216 96315: gap of      100 bp
          *    96316   109238: contig of 12923 bp in length
          *   109239 109338: gap of      100 bp
          *   109339   129556: contig of 20218 bp in length
          *   129557 129656: gap of      100 bp
          *   129657   186598: contig of 56942 bp in length
          *   186599 186698: gap of      100 bp
          *   186699   189025: contig of 2327 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189025
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-705C3"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .978
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1079. .1991
                     /note="assembly_fragment"
     misc_feature    2092. .4062
                     /note="assembly_fragment"
     misc_feature    4163. .77746
                     /note="assembly_fragment"
     misc_feature    77847. .83726
                     /note="assembly_fragment"
     misc_feature    83827. .96215
                     /note="assembly_fragment"
     misc_feature    96316. .109238
                     /note="assembly_fragment"
     misc_feature    109339. .129556
                     /note="assembly_fragment"
     misc_feature    129657. .186598
                     /note="assembly_fragment"
     misc_feature    186699. .189025
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    57529 a  38428 c  37721 g  54447 t    900 others
```

```
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 2;  Length 189025;
  Best Local Similarity   87.5%;  Pred. No. 1.8e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcgt 25
           |||| ||||||| ||||||||| |
Db 174190 CATCACCAGCATCTCGATCCTCCT 174213


RESULT  23
AC011276/c
LOCUS       AC011276    80954 bp    DNA             HTG       06-DEC-2000
DEFINITION  Homo sapiens clone RP11-13J17, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   AC011276
VERSION     AC011276.4  GI:11560229
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 80954)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome, clone RP11-13J17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 80954)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 6, 2000 this sequence version replaced gi:7230084.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
                Center project name: L3339
                Center clone name: 13_J_17
            ---------- Summary Statistics
                Assembly program: Phrap; version 0.960731
                Consensus quality:  bases at least Q40
                Consensus quality:  bases at least Q30
                Consensus quality:  bases at least Q20
                Insert size: 170000; agarose-fp
                Insert size: ; sum-of-contigs
                Quality coverage: 0.0 in Q20 bases; agarose-fp
                Quality coverage: 0.0 in Q20 bases; sum-of-contigs
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
```

```
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    19224: contig of 19224 bp in length
            *    19225 19324: gap of      100 bp
            *    19325    22150: contig of 2826 bp in length
            *    22151 22250: gap of      100 bp
            *    22251    27044: contig of 4794 bp in length
            *    27045 27144: gap of      100 bp
            *    27145    33158: contig of 6014 bp in length
            *    33159 33258: gap of      100 bp
            *    33259    53706: contig of 20448 bp in length
            *    53707 53806: gap of      100 bp
            *    53807    74158: contig of 20352 bp in length
            *    74159 74258: gap of      100 bp
            *    74259    80954: contig of 6696 bp in length.
FEATURES             Location/Qualifiers
     source          1. .80954
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-13J17"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    21767 a  18158 c  17386 g  23043 t    600 others
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 2;  Length 80954;
  Best Local Similarity   84.0%;  Pred. No. 3.5e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           | |||||||||| | ||||||| ||
Db   29333 GGATCGCCAGCAGTGCGATCCTAGT 29309


RESULT  24
AC068024/c
LOCUS       AC068024   112808 bp    DNA             HTG       30-MAY-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-802H23 map 17, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC068024
VERSION     AC068024.2  GI:8102491
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112808)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-802H23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 112808)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.

  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 30, 2000 this sequence version replaced gi:7657812.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10097
                Center clone name: 802_H_23
            --------------.
            * NOTE: This record contains 132 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      809: contig of 809 bp in length
            *      810 909: gap of      100 bp
            *      910     1714: contig of 805 bp in length
            *     1715 1814: gap of      100 bp
            *     1815     2645: contig of 831 bp in length
            *     2646 2745: gap of      100 bp
            *     2746     3565: contig of 820 bp in length
            *     3566 3665: gap of      100 bp
            *     3666     4510: contig of 845 bp in length
            *     4511 4610: gap of      100 bp
            *     4611     5441: contig of 831 bp in length
            *     5442 5541: gap of      100 bp
            *     5542     6334: contig of 793 bp in length
            *     6335 6434: gap of      100 bp
            *     6435     7241: contig of 807 bp in length
            *     7242 7341: gap of      100 bp
            *     7342     8132: contig of 791 bp in length
            *     8133 8232: gap of      100 bp
            *     8233     8954: contig of 722 bp in length
            *     8955 9054: gap of      100 bp
            *     9055     9864: contig of 810 bp in length
            *     9865 9964: gap of      100 bp
            *     9965    10772: contig of 808 bp in length
            *    10773 10872: gap of      100 bp
            *    10873    11658: contig of 786 bp in length
            *    11659 11758: gap of      100 bp
            *    11759    12546: contig of 788 bp in length
            *    12547 12646: gap of      100 bp
            *    12647    13470: contig of 824 bp in length
            *    13471 13570: gap of      100 bp
            *    13571    14380: contig of 810 bp in length
            *    14381 14480: gap of      100 bp
            *    14481    15286: contig of 806 bp in length
            *    15287 15386: gap of      100 bp
            *    15387    16188: contig of 802 bp in length
            *    16189 16288: gap of      100 bp
            *    16289    17102: contig of 814 bp in length
            *    17103 17202: gap of      100 bp
            *    17203    18046: contig of 844 bp in length
            *    18047 18146: gap of      100 bp
            *    18147    18952: contig of 806 bp in length
            *    18953 19052: gap of      100 bp
            *    19053    19876: contig of 824 bp in length
            *    19877 19976: gap of      100 bp
            *    19977    20778: contig of 802 bp in length
            *    20779 20878: gap of      100 bp
```

```
*    20879    21692: contig of 814 bp in length
*    21693 21792: gap of      100 bp
*    21793    22622: contig of 830 bp in length
*    22623 22722: gap of      100 bp
*    22723    23542: contig of 820 bp in length
*    23543 23642: gap of      100 bp
*    23643    24477: contig of 835 bp in length
*    24478 24577: gap of      100 bp
*    24578    25389: contig of 812 bp in length
*    25390 25489: gap of      100 bp
*    25490    26306: contig of 817 bp in length
*    26307 26406: gap of      100 bp
*    26407    27210: contig of 804 bp in length
*    27211 27310: gap of      100 bp
*    27311    28129: contig of 819 bp in length
*    28130 28229: gap of      100 bp
*    28230    29038: contig of 809 bp in length
*    29039 29138: gap of      100 bp
*    29139    29723: contig of 585 bp in length
*    29724 29823: gap of      100 bp
*    29824    30631: contig of 808 bp in length
*    30632 30731: gap of      100 bp
*    30732    31532: contig of 801 bp in length
*    31533 31632: gap of      100 bp
*    31633    32447: contig of 815 bp in length
*    32448 32547: gap of      100 bp
*    32548    33347: contig of 800 bp in length
*    33348 33447: gap of      100 bp
*    33448    34246: contig of 799 bp in length
*    34247 34346: gap of      100 bp
*    34347    35150: contig of 804 bp in length
*    35151 35250: gap of      100 bp
*    35251    36035: contig of 785 bp in length
*    36036 36135: gap of      100 bp
*    36136    36941: contig of 806 bp in length
*    36942 37041: gap of      100 bp
*    37042    37859: contig of 818 bp in length
*    37860 37959: gap of      100 bp
*    37960    38778: contig of 819 bp in length
*    38779 38878: gap of      100 bp
*    38879    39686: contig of 808 bp in length
*    39687 39786: gap of      100 bp
*    39787    40616: contig of 830 bp in length
*    40617 40716: gap of      100 bp
*    40717    41545: contig of 829 bp in length
*    41546 41645: gap of      100 bp
*    41646    42457: contig of 812 bp in length
*    42458 42557: gap of      100 bp
*    42558    43379: contig of 822 bp in length
*    43380 43479: gap of      100 bp
*    43480    44279: contig of 800 bp in length
*    44280 44379: gap of      100 bp
*    44380    45193: contig of 814 bp in length
*    45194 45293: gap of      100 bp
*    45294    46089: contig of 796 bp in length
*    46090 46189: gap of      100 bp
*    46190    46999: contig of 810 bp in length
*    47000 47099: gap of      100 bp
*    47100    47914: contig of 815 bp in length
*    47915 48014: gap of      100 bp
*    48015    48820: contig of 806 bp in length
*    48821 48920: gap of      100 bp
*    48921    49629: contig of 709 bp in length
*    49630 49729: gap of      100 bp
*    49730    50532: contig of 803 bp in length
*    50533 50632: gap of      100 bp
*    50633    51445: contig of 813 bp in length
*    51446 51545: gap of      100 bp
*    51546    52345: contig of 800 bp in length
*    52346 52445: gap of      100 bp
*    52446    53246: contig of 801 bp in length
*    53247 53346: gap of      100 bp
*    53347    54170: contig of 824 bp in length
*    54171 54270: gap of      100 bp
*    54271    55062: contig of 792 bp in length
*    55063 55162: gap of      100 bp
*    55163    55945: contig of 783 bp in length
*    55946 56045: gap of      100 bp
*    56046    56865: contig of 820 bp in length
*    56866 56965: gap of      100 bp
*    56966    57778: contig of 813 bp in length
*    57779 57878: gap of      100 bp
*    57879    58700: contig of 822 bp in length
*    58701 58800: gap of      100 bp
*    58801    59639: contig of 839 bp in length
*    59640 59739: gap of      100 bp
*    59740    60562: contig of 823 bp in length
*    60563 60662: gap of      100 bp
*    60663    61474: contig of 812 bp in length
*    61475 61574: gap of      100 bp
*    61575    62379: contig of 805 bp in length

  Query Match             74.4%;  Score 18.6;  DB 2;  Length 112808;
  Best Local Similarity   84.0%;  Pred. No. 3.5e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||| ||||||||| | ||||| |||
Db   42421 GCATGGCCAGCATGTTGATCCCCGT 42397


RESULT  25
AC055855/c
LOCUS       AC055855   159738 bp    DNA             HTG       04-JUL-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-352G18 map 15, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC055855
VERSION     AC055855.5  GI:14595940
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159738)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-352G18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 159738)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 4, 2001 this sequence version replaced gi:13959302.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
```

```
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9770
                Center clone name: 352_G_18
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 37% of reads
                Sequencing vector: Plasmid; n/a; 63% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 155836 bases at least Q40
                Consensus quality: 157373 bases at least Q30
                Consensus quality: 158080 bases at least Q20
                Insert size: 149000; agarose-fp
                Insert size: 158938; sum-of-contigs
                Quality coverage: 10.4 in Q20 bases; agarose-fp
                Quality coverage: 9.7 in Q20.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    15717: contig of 15717 bp in length
            *    15718 15817: gap of      100 bp
            *    15818    23190: contig of 7373 bp in length
            *    23191 23290: gap of      100 bp
            *    23291    32050: contig of 8760 bp in length
            *    32051 32150: gap of      100 bp
            *    32151    46572: contig of 14422 bp in length
            *    46573 46672: gap of      100 bp
            *    46673    59364: contig of 12692 bp in length
            *    59365 59464: gap of      100 bp
            *    59465    90204: contig of 30740 bp in length
            *    90205 90304: gap of      100 bp
            *    90305   112333: contig of 22029 bp in length
            *   112334 112433: gap of      100 bp
            *   112434   149286: contig of 36853 bp in length
            *   149287 149386: gap of      100 bp
            *   149387   159738: contig of 10352 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159738
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-352G18"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .15717
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    15818. .23190
                     /note="assembly_fragment"
     misc_feature    23291. .32050
                     /note="assembly_fragment"
     misc_feature    32151. .46572
                     /note="assembly_fragment"
     misc_feature    46673. .59364
                     /note="assembly_fragment"
     misc_feature    59465. .90204
                     /note="assembly_fragment"
     misc_feature    90305. .112333
                     /note="assembly_fragment"
     misc_feature    112434. .149286
                     /note="assembly_fragment"
     misc_feature    149387. .159738
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    43978 a  35268 c  35918 g  43757 t    817 others
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 2;  Length 159738;
  Best Local Similarity   84.0%;  Pred. No. 3.5e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           | |||||||||| | ||||||| ||
Db   25419 GGATCGCCAGCAGTGCGATCCTAGT 25395


RESULT  26
EHBIOTOP2
LOCUS       EHBIOTOP2    1041 bp    DNA             BCT       04-NOV-1996
DEFINITION  Erwinia herbicola biotin synthetase (bioB) gene, complete cds.
ACCESSION   U38648
VERSION     U38648.1  GI:1228112
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Erwinia herbicola strain=Eho10.
  ORGANISM  Pantoea agglomerans
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Pantoea.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Wu,C.H., Chen,H.Y. and Shiuan,D.
  TITLE     Isolation and characterization of the Erwinia herbicola bio operon
            and the sequences of the bioA and bioB genes
  JOURNAL   Gene 174 (2), 251-258 (1996)
  MEDLINE   97045821
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Shiuan,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-1995) David Shiuan. Biology, Natl. Sun Yat-Sen
            University, Kaohsiung, Taiwan, ROC
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Pantoea agglomerans"
                     /strain="Eho10"
                     /db_xref="taxon:549"
     gene            1. .1041
                     /gene="bioB"
     CDS             1. .1041
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAC44534.1"
                     /db_xref="GI:1228114"
                     /translation="MAHHARWTLSQVTELFDKPLLDLLFEAQTIHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLETERLMEVEQVLDSARKAKNAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLDETQAQRLASAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLDKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSFVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDVQLFRKLGLNPQQTDVMTGDNEQQQKLEEQIFN
                     ADTDQFYNAAAL"
BASE COUNT      230 a    322 c    308 g    181 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 1;  Length 1041;
  Best Local Similarity   87.0%;  Pred. No. 5.8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcg 24
           ||||| ||||| ||||| |||||
Db      91 CATCGTCAGCACTTCGACCCTCG 113


RESULT  27
```

```
D88424/c
LOCUS       D88424       3291 bp    mRNA            PLN       07-FEB-1999
DEFINITION  Cyanidium caldarium mRNA for plasma membrane H+-ATPase, complete
            cds.
ACCESSION   D88424
VERSION     D88424.1  GI:2208934
KEYWORDS    plasma membrane H+-ATPase.
SOURCE      Cyanidium caldarium cDNA to mRNA.
  ORGANISM  Cyanidium caldarium
            Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales;
            Porphyridiaceae; Cyanidium.
REFERENCE   1  (bases 1 to 3291)
  AUTHORS   Ohta,H., Shirakawa,H., Uchida,K., Yoshida,M., Matuo,Y. and Enami,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1996) to the DDBJ/EMBL/GenBank databases.
            Hisataka Ohta, Science University of Tokyo, Department of Biology;
            1-3 Kagurazaka, Shinjuku, Tokyo 162, Japan
            (E-mail:ohta@rs.noda.sut.ac.jp, Tel:03-3260-4271, Fax:03-3260-0322)
REFERENCE   2  (sites)
  AUTHORS   Ohta,H., Shirakawa,H., Uchida,K., Yoshida,M., Matuo,Y. and Enami,I.
  TITLE     Cloning and sequencing of the gene encoding the plasma membrane
            H(+)-ATPase from an acidophilic red alga, Cyanidium caldarium
  JOURNAL   Biochim. Biophys. Acta 1319 (1), 9-13 (1997)
  MEDLINE   97261359
FEATURES             Location/Qualifiers
     source          1. .3291
                     /organism="Cyanidium caldarium"
                     /db_xref="taxon:2771"
     CDS             235. .3102
                     /codon_start=1
                     /product="plasma membrane H+-ATPase"
                     /protein_id="BAA20486.1"
                     /db_xref="GI:2208935"
                     /translation="MADVESGNSQVNGGNEVDGTNGSSSLRAESKPLPADLDKRSLSM
                     SMLETVDLEKDDMDYVMACLKTSPEGLKPDVAARRLAKFGPNALPEKKVNPILEFLMF
                     MWNPLSWVMEAAALVAIFLTIPGGKAPDWEDFLGILLLLLINSTIGFIEERNAGNAVK
                     ALMDALAPRAKVQRGGEWLDIDAKDLVIGDIVALKLGDVIPADARIMNGKDIKIDQAA
                     LTGESLPVGKEKGDMIYSGSVVKQGEFLALVIATGMNTFFGKAAHLVNQTESTSHLQA
                     IVSAIGLYCMAWISTFVLLLIVTQWPIHLENYRHGINNILVLLIGGVPIAMPVVLSVT
                     LAIGAHELAEQKAIVTRMTAVEELAGMTILCSDKTGTLTLNKLSIDQESFFTMGGYTV
                     DTVDQCMVFAARASRTENQDAIDFAVVNSLPDPKMAREGIEELDFHPFNPVDKRTEIT
                     YRDNKDGKVYKATKGAPQIILGMAHNKKEIEKEVHEQIEDFAKRGFRALGIAVAEVPS
                     GEAHGEPGPWSMVGLMPIFDPPRHDTKETIEQAIAMGVEVKMITGDQLAIAKETARRL
                     GMGTNIFNTDVLNLSDQRASIEYGGSVGELVESADGFAGVFPEHKYRIVEVLQRRGHM
                     VGMTGDGVNDAPALKRASVGIAVAGATDAARGASDIVLTEPGLSVIIHAMVMSRQIFQ
                     RMKNYSMYACSVTVRIVVTFSILVWAFRFNMPPFLVLILAYLNDGTIMTISKDRVKPS
                     PLPQRWDLKEVFIVASSLGIYLTASTVIFYVTLFKTQFWHDTFKLGMPWLNPRDPNYF
                     QLHSIIYLQASIIGQALIFVTRAHWFFFMDRPGILLMSAFVVAQLVATFICVYANWGF
                     TQIQGTGWGWAGVVWVWNVIWYAPLDIIKIAVRSIITGDKTPIHKLFAARRMFTFDYS
                     KHGREGRMPRSSLQAAQARASVHRSMETYRASLQKNVNSLD"
     polyA_site      3291
                     /note="45 a nucleotides"
BASE COUNT      755 a    814 c    929 g    793 t
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 8;  Length 3291;
  Best Local Similarity   87.0%;  Pred. No. 5.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctc 23
           |||| |||||||||| || |||||
Db     707 GCATTGCCAGCATTGCGTTCCTC 685
```

```
RESULT  28
E13998/c
LOCUS       E13998       3451 bp    DNA             PAT       24-JUN-1998
DEFINITION  Cyanidium caldarium mRNA for proton-translocating ATPase.
ACCESSION   E13998
VERSION     E13998.1  GI:3252765
KEYWORDS    JP 1997252786-A/1.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 3451)
  AUTHORS   Ota,N., Enami,I., Uchida,K. and Matsuo,T. .
  TITLE     PROTON-ATPASE GENE
  JOURNAL   Patent: JP 1997252786-A 1 30-SEP-1997;
            ORIENTAL YEAST CO LTD
COMMENT     OS   Cyanidium caldarium
            PN   JP 1997252786-A/1
            PD   30-SEP-1997
            PF   25-MAR-1996 JP 1996124124
            PI   OTA NAOTAKA, ENAMI ISAO, UCHIDA KOJI, MATSUO TAKESHI PC
            C12N15/09,C12N1/21,C12N9/14,(C12N15/09,C12R1:89),(C12N1/21, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .3451
            FT                   /organism='Cyanidium caldarium' FT   CDS
                  340. .3207
            FT                   /product='proton-translocating ATPase'.
FEATURES             Location/Qualifiers
     source          1. .3451
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      818 a    860 c    960 g    813 t
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 6;  Length 3451;
  Best Local Similarity   87.0%;  Pred. No. 5.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctc 23
           |||| |||||||||| || |||||
Db     812 GCATTGCCAGCATTGCGTTCCTC 790
```

```
RESULT  29
KPU31464/c
LOCUS       KPU31464     4304 bp    DNA             BCT       07-JAN-1997
DEFINITION  Klebsiella pneumoniae citrate fermentation regulatory genes: sensor
            kinase CitA (citA), response regulator CitB (citB) and
            transcriptional regulator LtrA (ltrA) genes, complete cds.
ACCESSION   U31464
VERSION     U31464.1  GI:924990
KEYWORDS    .
SOURCE      Klebsiella pneumoniae.
  ORGANISM  Klebsiella pneumoniae
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Klebsiella.
REFERENCE   1  (bases 1 to 4304)
  AUTHORS   Bott,M., Meyer,M. and Dimroth,P.
  TITLE     Regulation of anaerobic citrate metabolism in Klebsiella pneumoniae
  JOURNAL   Mol. Microbiol. 18 (3), 533-546 (1995)
  MEDLINE   96342382
REFERENCE   2  (bases 1 to 4304)
  AUTHORS   Bott,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-1995) Michael Bott, Mikrobiologisches Institut,
            Eidgenoessische Technische Hochschule Zuerich, Schmelzbergstrasse
            7, Zuerich CH-8092, Switzerland
FEATURES             Location/Qualifiers
     source          1. .4304
                     /organism="Klebsiella pneumoniae"
                     /strain="ATCC 13882"
                     /db_xref="taxon:573"
     CDS             157. .1800
                     /gene="citA"
                     /note="sensor kinase of a two-component regulatory system;
                     essential for expression of citrate-specific fermentation
                     genes in Klebsiella pneumoniae;  histidine autokinase"
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="sensor kinase CitA"
                     /protein_id="AAC44733.1"
                     /db_xref="GI:924991"
                     /translation="MSIYPMYTRKITHWFARRSFQNRIFLLILFTSTIVMLAMSWYLT
                     DITEERLHYQVGQRALIQAMQISAMPELVEAVQKRDLARIKALIDPMRSFSDATYITV
                     GDASGQRLYHVNPDEIGKSMEGGDSDEALINAKSYVSVRKGSLGSSLRGKSPIQDATG
                     KVIGIVSVGYTIEQLENWLSLQISSLLIPMAIMLLLLLFCARRFSLHIKKQMLNMEPQ
                     QLSQLLIQQSVLFESVFEGLIAIDSDYKITAINQTARRLLNLSQPEPTLIGKRISSVI
                     SQEVFFYDAPQTNKKDEIVTFNQIKVIASRMAVILNNEPQGWVISFRSKDDINTLSLQ
                     LSQVQQYADNLRAVQHEHRNLISTIAGLLFLKRYNQALELIQQQSESHQKVIDFIARN
                     FQDNHLAGLLIGKYYRAKELGLELIFDPACFVDRLPTALSHNEWISIVGNLLDNAYNA
                     SLRQPQGSKQIECLINSDGQEVIIEIADQGCGIDEALRDRIFERGVTSSASKDHGIGL
                     WLVRSYVEQAGGSIVVENNIPFGTIFTLYIPLTRDEHHG"
misc_feature         226. .288
                     /gene="citA"
                     /note="encodes putative transmembrane helix which with the
                     other putative transmembrane helix region encloses a
                     presumably periplasmic domain"
gene                 226. .754
                     /gene="citA"
misc_feature         688. .754
                     /gene="citA"
                     /note="encodes putative transmembrane helix which with the
                     other putative transmembrane helix region encloses a
                     presumably periplasmic domain"
misc_feature         1204. .1206
                     /gene="citA"
                     /note="encodes autophosphorylation site, presumably at His
                     350"
gene                 1793. .2497
                     /gene="citB"
CDS                  1793. .2497
                     /gene="citB"
                     /note="response regulator of a two-component regulatory
                     system; transcriptional regulator essential for expression
                     of citrate-specific fermentation genes in Klebsiella
                     pneumoniae;  response regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="response regulator CitB"
                     /protein_id="AAC44734.1"
                     /db_xref="GI:924992"
                     /translation="MDSITTLIVEDEPMLAEILVDNIKQFPQFDVIGIADKLESARKQ
                     LRLYQPQLILLDNFLPDGKGIDLIRHAVSTHYKGRIIFITADNHMETISEALRLGVFD
                     YLIKPVHYQRLQHTLERFARYRSSLRSSEQASQLHVDALFNIQAREQTEPASAPLRGI
                     DESTFQRVLQLFADPTVVHTADSLARILGSSKTTARRYLEQGVKNDFLEAEISYGKVG
                     RPERIYHGKQTYPEQR"
misc_feature         1958. .1960
                     /gene="citB"
                     /note="encodes putative residue (Asp 56) phosphorylated by
                     phospho-CitA"
misc_feature         2333. .2392
                     /gene="citB"
                     /note="encodes putative DNA-binding helix-turn-helix
                     motif"
gene                 3225. .4154
                     /gene="ltrA"
CDS                  3225. .4154
                     /gene="ltrA"
                     /note="target genes regulated by LtrA are unknown;
                     LysR-type regulator A;  LysR-type transcriptional
                     regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator LtrA"
                     /protein_id="AAC44735.1"
                     /db_xref="GI:924993"
                     /translation="MNLNLLPDLALFVQIVDQGSFSAVARQNGITPSAVSRSVSRLER
                     EMGCKLLQRTTRKLRLSDAGETIYQHAQQMLEAARQAMDSACSRQTVAQGKLTLSVPK
                     AVGRFVIHPLMMAFFHRYPQVDVCLRLEDRPLDFIDDGIDLALRITDTPSPGLHGKPL
                     MPIRHVICATEAYLQQHGTPYTPQDLRAHSCISLGETPADARWKFRREGKTETVQTYG
                     RYAANHTAVRLDAVRQHLGIGSLPLFTAREALANGDIVQVLPEWEFISSYSGDLWLLW
                     AGDKHMPARMRAMIDYLSETVPALNAGSTEPAK"
misc_feature         3285. .3344
                     /gene="ltrA"
                     /note="encodes putative DNA-binding helix-turn-helix
                     motif"
BASE COUNT       995 a   1277 c   1145 g    887 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 1;  Length 4304;
  Best Local Similarity   87.0%;  Pred. No. 5.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcg 24
           ||||||||||||  ||||| |||
Db     273 CATCGCCAGCATGACGATCGTCG 251


RESULT  30
AE004457
LOCUS       AE004457    10239 bp    DNA             BCT       30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 18 of 529 of the complete
            genome.
ACCESSION   AE004457 AE004091
VERSION     AE004457.1  GI:9946024
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 10239)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 10239)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .10239
                     /organism="Pseudomonas aeruginosa"
                     /strain="PA01"
                     /db_xref="taxon:287"
     gene            155. .880
                     /gene="PA0190"
     CDS             155. .880
                     /gene="PA0190"
                     /codon_start=1
                     /transl_table=11
                     /product="probable acid phosphatase"
                     /protein_id="AAG03580.1"
                     /db_xref="GI:9946025"
                     /translation="MNNKTLCPSLLLCLSLLAPLSCLGETAAAPYPLAHPPRLADYLP
                     PPPAADSAAAVADLGAVLEAQRLRTPEQVRRVRAHDQWEDNVFPFAGDLLGASFDKER
                     LPLTRSFFNRAQENLVEVLMPAKKHFARPRPYEVTPKVKPVLPPPEGESYPSGHTMDS
                     YFKASLLSMLVPEHHDAFFARAEEHAQSRVLAGVHFPSDLEGGQTAAAALVASLLADP
                     AVAADFAAVREELRGALGLPKLQ"
     gene            complement(904. .1821)
```

```
                /gene="PA0191"
     CDS        complement(904. .1821)
                /gene="PA0191"
                /codon_start=1
                /transl_table=11
                /product="probable transcriptional regulator"
                /protein_id="AAG03581.1"
                /db_xref="GI:9946026"
                /translation="MDLRQLRYFIALAEHRSFVRAADAMGITQPAFSRSIQGLEQELG
                CQLVDRGSKDLRPTPEGQVVLQHALSLVQGSANLIHEIAQLNKLDAGELRFGSGPAPA
                QQLVPDAVADFINRHPRVQISLDVDNWEKLSRALLREEIEFFVADIRDFEADPNFQTR
                PLTPRRGQFFCRPQHPLLGKDSLSTNDLFNYPLASTLIPPGIRKLLANLSGKIDFAAN
                IRCENMHALIRIVRQTDAIGIASDETLRPYLQRGELQPLQWRNLPQNLDSLSARCGII
                SRSGYRLSAAAKAMIETLVELDQAQHAAA"
     gene       2171. .4543
                /gene="PA0192"
     CDS        2171. .4543
                /gene="PA0192"
                /codon_start=1
                /transl_table=11
                /product="probable TonB-dependent receptor"
                /protein_id="AAG03582.1"
                /db_xref="GI:9946027"
                /translation="MKHHYRLPLRRTPPGGSDYALALLLALGGALPASAADAPAPAEE
                APPLAGSAPAKADTALGKVTVTARRREEDSQKVPTPITVLGGETLEAQRISRVQDLQQ
                VLPSVNVAYIHARQSSVAVRGIGNNPASDGLEGSAGIYLDNVYLGRPGMAVFDLLDIE
                QLELLRGPQGTLFGKNTTAGVLNISTRAPTFTAERTVEVSGGQDGYFQGRGTVSGPLG
                ETLAGRLSAYRTRDDGYIKNIHDDNYLNGGERQGARGQLLFEPNEDFSLRWIADYNEE
                DSSNGSMVVYGGAERFWQRAALVGASPLRDPQRRKVNINGRQHVSVHQGGSSLEANWN
                LAGGYRLTSISAYRYWHFTPANDEQLNVSAINDTGVEVHDRQFSQEIRLASPTGGAFD
                YVVGAYAFRQNLGNKTFTSYGPLADLYLLGANLGALNDTYSKANGKIETDSFALFAQG
                TWHLTERLDFTAGLRGTYEEKNAKVERFAPLGGAAVGGVGAAVRNGQLGAYDSGDLSQ
                YNFAPSALLSLSYQFSDDLLGYASLSHGEKSGGVNLAVGSAPSAGADSLLVGPERAND
                AELGLKSTLFDRRLLLNANLFWTGIHGYQATTLYQAPGSTQLVQVLANAGSVRSRGLE
                FEATALPLRGLTLNFNGSYNDVTYLSFKDAPCPAEVSTRPGAPSSCDLTGQRVVGASK
                WIANLNGEYQWRLDDRFQPYVSASYAYRSAAEGTLDNSDLSKIDGYALVNLAAGLRSD
                LGDGQLDTSVWLKNAFDKDYYLSAFASINGSYTASVGQPRTLGVSLRYDF"
     gene       4584. .5486
                /gene="PA0193"
     CDS        4584. .5486
                /gene="PA0193"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG03583.1"
                /db_xref="GI:9946028"
                /translation="MSNAAQALATPVRLDIRPVAGRIGAEIRGVTLSGELDAATVEAI
                QAALVRHKVIFFRDQSHLDDQTQEAFAHLLGEPVAHPTVPSREGTRFLLELDGAEGRR
                ANSWHTDVTFVEAYPKASILRSVVAPESGGDTVWANTASAYADLPAELRELADRLWAV
                HSNEYDYAGVKPSASVEQLENYRKVFTSTVYETEHPVVRVHPQSGERTLLLGHFVKRL
                KGLSQHDSAHLFAVLQGHVTRLENTVRWRWQAGDVAIWDNRATQHYAIDDYGDQPRVV
                RRVTLGGDVPVGIDGQRSRTIRKA"
     gene       5554. .6453
                /gene="PA0194"
     CDS        5554. .6453
                /gene="PA0194"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG03584.1"
                /db_xref="GI:9946029"
                /translation="MSQSATARQPEPEVAEAFRITPLEAPLGAEVRGLDARRPLAPEQ
                VLALKQALREHHILVFRQQHLDDEQYLRFATLFGSVFQPPADIPVLSSGGDGKVPDIV
                KVANTGDGELGNFALPAHIDHQWTPVPSSGSFLYALEVPSSGGETRFTNLARAYESLD
                EATRREIDGLRLINYNPFIRLREGGYGGGFATYRTPDIEPIQGSEHPLVRTHPESGRR
                VLFLSAHTEVEIPGYDPARGQALIGRLREHLARPELSYSHAWSVGDIVWWDNQAVLHA
                RNAFPASERRRLKRISLAGSRPF"
     gene       7100. .8602
                /gene="pntA"
                /note="PA0195; PA0195 - This gene contains a frameshift
                that has been confirmed in the final assembly. Contains
                high homology to C-terminus of pntA"
     gene       8602. .10038
                /gene="pntB"
                /note="PA0196"
     CDS        8602. .10038
                /gene="pntB"
                /codon_start=1
                /transl_table=11
                /product="pyridine nucleotide transhydrogenase, beta
                subunit"
                /protein_id="AAG03585.1"
                /db_xref="GI:9946030"
                /translation="MSMNLVTLLYLIASVCFIQALKGLSHPTTSRRGNLFGMVGMAIA
                VATTVGLVFKLGAEIATTGVGYIVVGLLVGGTAGSIMAKRVEMTKMPELVAFMHSMIG
                LAAVFIAIAAVVEPQSLGIVAHLGDTIPTGNRLELFLGAAIGAITFSGSVIAFGKLSG
                KYKFRLFQGTPVQFSGQHLLNLVLGLATLGLGLVFMFTGNLTAFAVMLALAFVLGVLI
                IIPIGGADMPVVVSMLNSYSGWAAAGIGFSLNNSMLIIAGSLVGSSGAILSYIMCKAM
                NRSFFNVILGGFGAEADAGGPAGSKEQRPVKSGSADDASFLLTNADSVIIVPGYGLAV
                ARAQHALMELAEKLTHRGVTVKFAIHPVAGRMPGHMNVLLAEAEVPYEQVFEMEDINS
                EFGQTDVVLVLGANDVVNPAAKNDPKSPIAGMPILEAYKAKTVIVNKRSMASGYAGLD
                NELFYLDKTMMVFGDAKKVIEDMVKAVE"
BASE COUNT      1705 a   3582 c   3290 g   1662 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 1;  Length 10239;
  Best Local Similarity   87.0%;  Pred. No. 5.6e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcg 24
           ||||||| ||||  |||||||||
Db    9861 CATCGCCGGCATGCCGATCCTCG 9883


RESULT  31
RLE310723
LOCUS       RLE310723   11382 bp    DNA             BCT       11-MAR-2001
DEFINITION  Rhizobium leguminosarum tonB gene, hmu operon (hmuPV genes) and
            rpoZ gene.
ACCESSION   AJ310723
VERSION     AJ310723.1  GI:13277325
KEYWORDS    hemin uptake protein; HmuP gene; HmuS gene; HmuT gene; HmuU gene;
            HmuV gene; ORF1; ORFA; ORFB; ORFC; ORFD; ORFE; rpoZ gene; sigma
            factor; tcrX gene; tcrY gene; tonB gene; two component regulator.
SOURCE      Rhizobium leguminosarum.
  ORGANISM  Rhizobium leguminosarum
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 11382)
  AUTHORS   Wexler,M., Yeoman,K.H., Stevens,J.B., De Luca,N.G., Sawers,G. and
            Johnston,A.W.B.
  TITLE     Rhizobium leguminosarum tonB is required for siderophore and haem
            uptake and is next to hmu haem uptake genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11382)
  AUTHORS   Wexler,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2001) Wexler M., Biological Sciences, University
            of East Anglia, Earlham rd Norwich, Norfolk, NR4 7TJ, UNITED
            KINGDOM
FEATURES             Location/Qualifiers
     source          1. .11382
                     /organism="Rhizobium leguminosarum"
                     /strain="8401(pRL1JI)"
                     /sub_species="biovar viciae"
                     /db_xref="taxon:384"
                     /country="United Kingdom"
     CDS             complement(1. .315)
                     /function="unknown"
                     /note="ORFA"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC34386.1"
                     /db_xref="GI:13277326"
```

```
                     /translation="MYIAMNRFKVATGSEGDFETVWRNRDSSLPEVPGFVEFRLLRGK
                     VNDEEGYTLYSSHTVWKSEADFQNWTKSMSFRAAHRNAGDHKAIYKGPPVFEGFNVVD
                     GI"
     mat_peptide     complement(4. .312)
                     /note="ORFA"
                     /product="hypothetical protein"
     CDS             complement(325. .597)
                     /function="unknown"
                     /note="ORFB"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC34387.1"
                     /db_xref="GI:13277327"
                     /translation="MNDQRPDGFTM1GLHKLAAQTGEGLVPELYELFQQHAERQQIYQ
                     NVMFFPTWEARMPGEVATRPLGPAAANGDNVLAFPSRAARAGKRKA"
     mat_peptide     complement(328. .594)
                     /note="ORFB"
                     /product="hypothetical protein"
     CDS             complement(594. .1001)
                     /function="unknown"
                     /note="ORFC"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC34388.1"
                     /db_xref="GI:13277328"
                     /translation="MARKGKKGSNREVRDSAGEDAGQASLGRSKLDGRSFLYVGGRDC
                     QVAHLRQICSNFGAELIHHDGGLREAVSRIDTVLPSVDCVFCPIDCISHDACLRVKTG
                     CKKFGKAFIPVRNGSKSSLERALQTMNERDNSR"
     mat_peptide     complement(597. .998)
                     /note="ORFC"
                     /product="hypothetical protein"
     gene            complement(1109. .2473)
                     /gene="tonB"
     CDS             complement(1109. .2473)
                     /gene="tonB"
                     /function="membrane protein; transduces energy to outer
                     membrane receptors"
                     /note="size of TonB protein not experimentally determined;
                     other possible start sites are at 2431 (440 residues) or
                     2158 (349 residues)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="TonB protein"
                     /protein_id="CAC34389.1"
                     /db_xref="GI:13277329"
                     /translation="MNPGGHGQTVRGIQMAISAKSRSRQVLIGEPDADGSLNDNNMHP
                     GHELSDLRNVQRQPAGEAVVHYARFAQIPAFPDHPEAEPIASVPAPPIDAAVEKQEDE
                     KKPVRRRVALTCIGSFFFHAGLLVALSVFMPTPPNEAIEDAGEAVSVVVYGDSDVDQT
                     SAGDPELQRQPEQVASEEVEPDTVQSEEATELNATTVPPEQAQPVETETMETVQSVQE
                     VTRVSPEAVVAAEPEVLVSESPAESFVAQPMATAVPEQPMPDIAQATVPEEVVPTAVQ
                     PTAVPPEEVKTVEVSPEPEDKPKPKAEKPKPVEKKQSQKRAKPAGGKEGSDREDSTRG
                     TVNGQHGPQTDGTSTTTGGTDGMGSAAVANYPGKIQKRIRRAVRVPDEYKNKSDGMTV
                     RIQLTINGTGRVASVSVARSSGIAELDKAVLDGVRRAAPFPPLPSEWGKPSWTFTQEV
                     QVTR"
     mat_peptide     complement(1112. .2470)
                     /gene="tonB"
                     /product="TonB"
     CDS             complement(2487. .2912)
                     /function="unknown"
                     /note="ORF1
                     orf1 is co-transcribed with tonB"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC34390.1"
                     /db_xref="GI:13277330"
                     /translation="MMTVRFAAVAAGVVMAVSSVCQAQDAATAKLDVELNALAPSQKG
                     CMMTFVAENNLQAPISKISFELAFFNDKNAVDRITVLDFRDLPQGKKRVRQFDMPNVK
                     CETVTRIIINDTPVCDGPATGECMKGLVTRSQISVPFEG"
     mat_peptide     complement(2490. .2909)
                     /note="ORF1"
                     /product="hypothetical protein"
     gene            3089. .3268
                     /gene="hmuP"
     CDS             3089. .3268
                     /gene="hmuP"
                     /function="involved in hemin uptake"
                     /codon_start=1
                     /transl_table=11
                     /product="HmuP hemin uptake protein"
                     /protein_id="CAC34391.1"
                     /db_xref="GI:13277331"
                     /translation="MVEKPDNFKHVPLQSEPAAQHRIVESADLFRGTNEIMIRHDGLV
                     YRLKITRQGKLILNK"
     mat_peptide     3092. .3265
                     /gene="hmuP"
                     /product="HmuP hemin uptake protein"
     gene            3277. .4326
                     /gene="hmuS"
     CDS             3277. .4326
                     /gene="hmuS"
                     /function="putative hemin degrading protein"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="HmuS protein"
                     /protein_id="CAC34392.1"
                     /db_xref="GI:13277332"
                     /translation="MTEQTRPAPAEIRTFRAENPKMRERDIAAQLKISEAALVAAETG
                     ISVTRIDGSALKLLERVAGLGEVMALSRNESAVHEKIGVFENIKSGAQAAIVLGENID
                     LRIFPSRWEHGFAVSKKDGDQERLSLQYFDKTGNAVHKVHLRPSSNVEAYHAIVAELK
                     LEDQSQEFVEAETSNAADDTADVSRDELRDNWSKLTDTHQFFGMLKRLKIGRQAAVRT
                     VGDDYAWRLDNSATAEMMHASVKSGLPIMCFVANDGIVQIHSGPIFNVQSMGPWINIM
                     DPTFHLHLRQDHIAETWAVRKPTTDGHVTSLEAYNAEGEMIIQFFGKRQEGSDERAAW
                     REIMENLPRAASVAA"
     mat_peptide     3280. .4323
                     /gene="hmuS"
                     /product="HmuS protein"
     gene            4338. .5270
                     /gene="hmuT"
     CDS             4338. .5270
                     /gene="hmuT"
                     /function="periplasmic hemin binding protein"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="HmuT protein"
                     /protein_id="CAC34393.1"
                     /db_xref="GI:13277333"
                     /translation="MTMRNNLRRIRTKAQLALTAAVMALPLIPTARPIEGFAFVRAAH
                     ADEKKLDTSRLVSVGGDITEIVYALGEESRLIARDTTSMYPEAALKLPNVGYMRALSP
                     EGILAMNPTAIIAVDGSGPQEALTVLKNASVPFETVPSAFTRDGIIAKIDRVGTLLGV
                     PDKAKALEEKVAADLDAAIADAEKRPEAERKRVLFILSAQNGRIMASGTGTAADGIVK
                     LAGAIKPVGAFPGYKPLTDEAIIEAKPDIILMMNRGDGAGTKNEDLLAQPAIVLTPAG
                     EKKAIIRMDGVYLLGFGPRTAAAARELNSAIYGG"
     mat_peptide     4341. .5267
                     /gene="hmuT"
                     /product="HmuT protein"
     gene            5273. .6385
                     /gene="hmuU"
     CDS             5273. .6385
                     /gene="hmuU"
                     /function="hemin permease"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="HmuU protein"
                     /protein_id="CAC34394.1"
                     /db_xref="GI:13277334"
                     /translation="MALPQAMAERRRPFPMADIREIRQAGDRTRLALLAIALLVAGSV
                     FSMLFSVTTGASDASILDVISNMAGSETALSTRDRIIIFDIRLPRAILGFLIGASLAV
                     SGTVMQGLFRNPLADPGLVGVSSGASLGAVAMIVLGSGLAAPLQALLGIYALPAAAFG
                     GGLVTTLLLYRIATRHGQTSVATMLLAGIALGALALAITGLLIYMANDQQLRDLTFWS
```

```
                     MGSLAGATWTKIAAASPIILLSFTALPFMARGLNAITLGEAAAFHMGVPVQRLKNVAI
                     VGVAAATGASVAVSGGIGFVGIVVPHILRMAIGPDHRFLLPAAALLGGSLLIFADVLA
                     RTLVAPAELPIGIITAAVGGPFFLWILLRQRSRLAL"
     mat_peptide     5276. .6382
                     /gene="hmuU"
                     /product="HmuU protein"
     gene            6422. .7216
                     /gene="hmuV"
     CDS             6422. .7216
                     /gene="hmuV"
                     /function="ATPase component of hemin transporter"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="HmuV protein"

  Query Match             72.8%;  Score 18.2;  DB 1;  Length 11382;
  Best Local Similarity   87.0%;  Pred. No. 5.6e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctc 23
           ||| |||| | ||||||||||||
Db   10445 GCAGCGCCGGGATTTCGATCCTC 10467


RESULT  32
AE004541/c
LOCUS       AE004541    11840 bp    DNA             BCT       30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 102 of 529 of the complete
            genome.
ACCESSION   AE004541 AE004091
VERSION     AE004541.1  GI:9947024
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 11840)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 11840)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .11840
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            complement(130. .1326)
                     /gene="PA1107"
     CDS             complement(130. .1326)
                     /gene="PA1107"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG04496.1"
                     /db_xref="GI:9947025"
                     /translation="MARTTLESIQHAIEVNSSLALPIALENLSRLTHLALLTVPFNLI
                     HILVFSLKDFRPDLGHQLWRQEIMYAHGAMALLFGGIGLLALWLRRQPPKLWRMRLLI
                     LLGGAGIIGFGVAIACIDQRITSNITPLLLACFACAMFILIRPAYAVPFYGLAMLAFE
                     VAMDHAQADPQLRLSNQANGLTAFGLGLLLSLILWHGHVRNLRQQRKLELQRQELEEK
                     NRQLEYLAGHDPLTGLFNRREFDQLVLMELARIARQPQPLSLLMVDLDHFKYINDRYG
                     HPFGDEVIRHAANLLRGNTRSSDSVARLGGEEFLLLLPDTAEQQARSLAEKLRERLEG
                     TPLATQDGVLCLTASFGVASLDGGGQASYEQLYSAADQALYRAKSSGRNRVEVIRLAS
                     LAESLR"
     gene            complement(1443. .2600)
                     /gene="PA1108"
     CDS             complement(1443. .2600)
                     /gene="PA1108"
                     /codon_start=1
                     /transl_table=11
                     /product="probable MFS transporter"
                     /protein_id="AAG04497.1"
                     /db_xref="GI:9947026"
                     /translation="MPDRPRPPLLLVLALLALPQVAETILSPALPALASHWRLDDATS
                     QWTMALFFVGFAPGIWLWGWLADRLGRRPALLGGLGLAALATFGAWASTDYSYLLACR
                     LVQGLGLATCSVTVQASLRDVLQGPALMSYFVTLGAVLAWSPAVGPLGGQWLADLGGH
                     PAVFATLAVLLASLAALVVPAWPETRPLLAGTPEPATLAIFRRVLADRPLQTRALLVA
                     VLNVLVFSFYAAGPFMVGDLPGLGFGWIGLAIAIAGSLGALLNRRLPRTWNSARRVRL
                     GLALAAAGATAQTLLAAVGYAEGLYWALPALPIFIGFGVAIPNLLGPALHAYDDCRGR
                     AGALFGLAYYLLIGLGLGASTLLPFDRPWPLAAYWSLLGLVALLLARRLPE"
     gene            2690. .3469
                     /gene="PA1109"
     CDS             2690. .3469
                     /gene="PA1109"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transcriptional regulator"
                     /protein_id="AAG04498.1"
                     /db_xref="GI:9947027"
                     /translation="MPRYPYERVEDPDLARCPVVGLEVDSTGHRSDWHCHRRAQLLYM
                     AAGSVTLYFAERICQLTPLQAAWLPAGVPHRTVLHGRFAYRSLYFEPSAFAELPDSVQ
                     VLEVGPLMRELILRMTGWAADLQLEGPHLTLLRALFDELASAPAHAASLPMPRDARAL
                     KVAEALLAQPALDWSIEAWGREVGASGRTLARLFTEQTGLGFSRWRTQARLFEARRRL
                     AEGRSVTEVAHALGYASDSAFIAMYRRVCGQSPGRALRGLE"
     gene            complement(3474. .4184)
                     /gene="PA1110"
     CDS             complement(3474. .4184)
                     /gene="PA1110"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG04499.1"
                     /db_xref="GI:9947028"
                     /translation="MTDPIRLSKRLAELTACSRREAELYIEGGWVSVDGEVIEEPQFK
                     VLDQRVELLPGARAETIEPATLLLHKPAGWRHDDFEGLLAAGRRWSDDPSPLRALKKH
                     FARQRPTLALDSEASGLVVFSQQHGVLRKLVDDGARLEQEYLVEVAGDLAAGGLERLR
                     HGLAYQGRRLSPCKVSWQNESHLRFALKDVLPGQLRFMCEREGLEVRSIRRLRIGALS
                     LARLPLGEWRYLGLHERF"
     gene            4297. .4920
                     /gene="PA1111"
     CDS             4297. .4920
                     /gene="PA1111"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG04500.1"
                     /db_xref="GI:9947029"
                     /translation="MQSRLIDGLEWDLDEDGSGGLCNVHGDKVHLRQGDMDGACGPYC
                     LVMAMLVRGQLRRRQAKGLERVDSRSRYGRLMKALDQHGPLLRVGTTGVDLLELLAEI
                     NDTEHFALKGEGRRLVRHTREYLEAGFPVVLGFHGRKGSDIRHWGLAIGMSERAFLLL
                     DPAHDLPRGQAWNAVITSEAHGSRFGYRYINAKGTWAVTLKEMVALL"
     gene            5013. .6161
                     /gene="PA1112"
     CDS             5013. .6161
                     /gene="PA1112"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG04501.1"
```

```
                     /db_xref="GI:9947030"
                     /translation="MPGLSLNAILLSAVALLAPSVEAVDAGRHDSELGAVEVTPVVGG
                     LAYPWAVAFLPDNQGILVTERPGSLRFVSPEGQLSAPLEGVPRVFASGQGGLLDVALS
                     PGFAEDRLVYLTYAEAGEDGRAGTAAGRGRLSADKRRLEGFHVIFRQLPKLSSGHHFG
                     SRLVFDGKGHLFIALGENNQRSTSQDLDKLQGKIVRLYPDGGVPKDNPFVGRADARPE
                     IWSYGHRNQQGAALNPWTGRLWTHEHGPRGGDEINIPQAGGNYGWPLATHGINYTLLP
                     IPEARGKAVEGTEAPHHVWEKSPAISGMAFYDAQRFPAWQHNLFIGALATQELIRLQL
                     DGDRIVHEERLLGDLGVRIRDVRQGPDGYLYLLTDQKDGQLLRLGLAR"
     gene            7151. .8917
     CDS             7151. .8917
                     /gene="PA1113"
                     /codon_start=1
                     /transl_table=11
                     /product="probable ATP-binding/permease fusion ABC
                     transporter"
                     /protein_id="AAG04502.1"
                     /db_xref="GI:9947031"
                     /translation="MPAILSSRQRIALRLAWRFMAPYRWRMLGALVALLFTAAITLSL
                     GQGIRLLVDQGFLTASPELLNRYILLFFVLVLGLAVGTFVRFYLVSWIGERFVADIRK
                     RVFDHLIELHPGFYETNRSSEIQSRLTADTTLLQTVIGSSLSMALRNALMLVGGVLLM
                     FVTNAKLTSIVVLALPLVIAPILLFGRRVRSLSRESQDRVADVGSYVGETLGQIKTVQ
                     AYNHQAQDRQRFAGTVERAFDTARKRIAQRAWLITVVIVLVLGAVGIMLWVGGRDVIA
                     GSISAGELAAFVFYSLIVGMAFGTLSEVIGELQRAAGAAERIAELLRARSEIRPPAQP
                     RKLAARVSGRIELQAVRFAYPTRPEHWAIDDISLAVEPGETLALVGPSGAGKSTLFDL
                     LLRFFDPQQGRLLLDGQPIVELDPADLRRSFALVSQNPALFFGSVEENIRYGRPHASA
                     AEVEAAARAAHAHEFILGLPQGYATHLGEGGIGLSGGQRQRLAIARALLVDAPVLLLD
                     EATSALDAESEHLIQQALPGLMSGRTTLVIAHRLATVQSADRIAVIDQGRLVAIGSHA
                     RLIAENPLYARLAELQFAAGGR"
     gene            complement(8920. .9255)
                     /gene="PA1114"
     CDS             complement(8920. .9255)
                     /gene="PA1114"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG04503.1"
                     /db_xref="GI:9947032"
                     /translation="MKHLAFLALGGLLLAPLAEAASFDCEKAERADEKAICANRALND
                     KDVEMSVKYRFLRGLFAMGGRGAMQDRQQDWLRERRRCGGDTDCLSRAYAQRIDELDK
                     QYAGIDKPL"
     gene            complement(9359. .11683)
                     /gene="PA1115"
     CDS             complement(9359. .11683)
                     /gene="PA1115"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG04504.1"
                     /db_xref="GI:9947033"
                     /translation="MRPLSRWLSSCGSLLALAALFLVVPLAGRYWLGWSTPFGYLSDL
                     ALGSLLVLLLHRRSLWFALPVMLLWCLFLIGNAELVAAVGRMPEPSDLKYLLDPQFVS
                     SSTQGGGLAHLPLALGLGGATLLCVLLWRRRGPRLPRYAYALPAAFLAAHVGLQYLRP
                     SDADQWQQYNLPHKLLASALSSGQLAVEDWLADDRPDTPPDVAGLNQLDLSGTPLLAG
                     GGRARNVLIVAMEGIPGAYIAANRAAINSSYHEAPMPRLSGWAERAMTTPDYVLHSHQ
                     TIRGLYAMLCGDYSKLDSGTPKGVELLNNPARARECLPAQLRQHGFSTHFLQGAGLRF
                     MAKDKVMPQMGFDKTLGRDWFRNTSYLEFPWGMDDKAFFEGAATYVKQLRQQKKPWML
                     TLLTVGTHQPYSAPADYLARYPSAKQAAVAYLDDAIDAFLADLEKQGVLRDTLVVLTS
                     DESHGVDQVRLASAWGFNLVLAPEQAALPPIKSGVYGHVDLAASVLDYFGYRVPAGIS
                     GRSLFRDYASGREMMSYTNGMLRHHDGKDTFVECDFQQVCRRYRSEGFIADEAEFLGR
                     FSGRQARLVSQRAAVLDHSVQQAGQLGQEFQFATHERIRLKPETHNDWTDNLIGAQYL
                     EMPKNSRTTVTLKIRALQMDRNGARVTLKTKEFERDVPIPVPDIPRLLKGKPLDMSFT
                     FDNPEARKAFSFHLLAEGRGIIEISDFHVVTAANPEPVESRSQVAEAAKEQPLLSTLK
                     SLFANRPETEEAERADAPRERTVIPEQKALILSH"
BASE COUNT     1889 a    3811 c    4174 g    1966 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 1;  Length 11840;
  Best Local Similarity   87.0%;  Pred. No. 5.6e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctc 23
           ||| |||| ||| ||||||||||
Db    8201 GCACCGCCTGCAATTCGATCCTC 8179


RESULT  33
ECOBIOB
LOCUS       ECOBIOB       128 bp    DNA             BCT       03-JAN-1995
DEFINITION  Escherichia coli biotin (bioB) gene, early terminator region.
ACCESSION   M27731
VERSION     M27731.1  GI:341755
KEYWORDS    bioB gene; biotin.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 128)
  AUTHORS   Nath,S.K.
  TITLE     Attenuation of transcription of biotin genes in Escherichia coli
  JOURNAL   Can. J. Microbiol. 34 (12), 1288-1296 (1988)
  MEDLINE   89167942
FEATURES             Location/Qualifiers
     source          1. .128
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     gene            4. .128
                     /gene="bioB"
     -10_signal      4. .9
                     /gene="bioB"
     mRNA            20. .>128
                     /gene="bioB"
     terminator      103. .121
                     /gene="bioB"
                     /note="early terminator"
BASE COUNT       31 a     28 c     34 g     35 t
ORIGIN


  Query Match             72.0%;  Score 18;  DB 1;  Length 128;
  Best Local Similarity   100.0%;  Pred. No. 7.4e+02;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 cagcatttcgatcctcgt 25
           ||||||||||||||||||
Db      57 CAGCATTTCGATCCTCGT 74


RESULT  34
MSU31280
LOCUS       MSU31280     3748 bp    DNA             BCT       21-NOV-1996
DEFINITION  Methylobacillus sp. beta-lactamase (bla) and biotin synthase (bioB)
            genes, complete cds.
ACCESSION   U31280
VERSION     U31280.1  GI:1680651
KEYWORDS    .
SOURCE      Methylobacillus sp. KT1.
  ORGANISM  Methylobacillus sp. KT1
            Bacteria; Proteobacteria; beta subdivision; Methylophilus group;
            Methylobacillus.
REFERENCE   1  (bases 1 to 3748)
  AUTHORS   Serebriiskii,I.G., Vassin,V.M. and Tsygankov,Y.D.
  TITLE     Two new members of the bio B superfamily: cloning, sequencing and
            expression of bio B genes of Methylobacillus flagellatum and
            Corynebacterium glutamicum
  JOURNAL   Gene 175 (1-2), 15-22 (1996)
  MEDLINE   97074643
REFERENCE   2  (bases 1 to 3748)
  AUTHORS   Serebrijski,I., Vassin,V. and Tsygankov,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-1995) Ilya Serebrijski, Institute of Genetics and
            Selection of Industrial Microorganisms, 1st Dorozhny Proezd,
            Moscow, 113545, Russia
FEATURES             Location/Qualifiers
```

```
     source          1. .3748
                     /organism="Methylobacillus sp. KT1"
                     /isolate="KT1"
                     /db_xref="taxon:79675"
     gene            201. .1061
                     /gene="bla"
     CDS             201. .1061
                     /gene="bla"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-lactamase"
                     /protein_id="AAC44581.1"
                     /db_xref="GI:1680652"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
     -35_signal      2141. .2153
                     /gene="bioB"
     gene            2141. .3243
                     /gene="bioB"
     -10_signal      2177. .2183
                     /gene="bioB"
     RBS             2207. .2211
                     /gene="bioB"
     CDS             2218. .3243
                     /gene="bioB"
                     /note="Method: conceptual translation supplied by author"
                     /codon_start=1
                     /transl_table=11
                     /function="sulfur inserting enzyme; last step of biotin
                     biosynthesis"
                     /product="biotin synthetase"
                     /protein_id="AAC44582.1"
                     /db_xref="GI:1680653"
                     /translation="MTMITNSHRMRDSVIDIRGLDRVQRARVQAKEEAPKRWSVDDIV
                     ALFELPFSDLMHRAQSVHRENFDPNGVQVSTLLSIKTGGCSEDCGYCPQAARYHTDVE
                     KQDLHAAGRGTRSRVLPRRMRQPFCMAAWQPKQRDLEPVLAMIREVKAMGLETCATLG
                     MLKDGQAEQLKEAGLDYYNHNLDTAPEYYGEVITTRTYQDRLDTLDRVREQDINVCCG
                     GIIGMGESRVQRAGLLAQLANMERPPESVPDQLLTQVEGTPMYGMDELDPFEFVRTIA
                     AARITMPQSFVRLSVGRQSMHEGIQALCFLAGANSIFYGEKLLTTGNPEAEADRKLFD
                     KLGIHPL"
BASE COUNT      895 a    980 c    988 g    885 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 1;  Length 3748;
  Best Local Similarity   90.5%;  Pred. No. 8.8e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcc 21
           ||||||| || ||||||||||
Db    2400 GCATCGCGAGAATTTCGATCC 2420


RESULT  35
CVU39574
LOCUS       CVU39574     3748 bp    DNA   circular  SYN       25-NOV-1995
DEFINITION  Expression cloning vector pBVI02, including biotin synthetase
            (bioB) and beta-lactamase (bla) genes, complete sequence.
ACCESSION   U39574
VERSION     U39574.1  GI:1073246
KEYWORDS    .
SOURCE      Cloning vector pBVI02.
  ORGANISM  Cloning vector pBVI02
            artificial sequence; vectors.
REFERENCE   1  (bases 1 to 3748)
  AUTHORS   Serebrijski,I., Vassin,V. and Tsygankov,Y.
  TITLE     Cloning, sequencing and expression of bioB genes of Methylobacillus
            flagellatum and Corynebacterium glutamicum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3748)
  AUTHORS   Serebrijski,I., Vassin,V. and Tsygankov,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-1995) Vitaly Vassin, Institute of Genetics and
            Selection of Industrial Microorganisms, 1st Dorozhny proezd, 1,
            Moscow, 113545, Russia
COMMENT     (U31280).
FEATURES             Location/Qualifiers
     source          1. .3748
                     /organism="Cloning vector pBVI02"
                     /db_xref="taxon:43991"
                     /note="bioB gene expression vector based on pUCBM20
                     replicon; bioB gene was placed under the control of the
                     Plac promoter; production of BioB protein in
                     E.coli[pBVI02] cells reaches up to 25% of the cell
                     protein"
     gene            201. .1061
                     /gene="bla"
     CDS             201. .1061
                     /gene="bla"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-lactamase"
                     /protein_id="AAA81772.1"
                     /db_xref="GI:1073247"
                     /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                     IELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVE
                     YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRL
                     DRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPL
                     LRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIA
                     EIGASLIKHW"
     promoter        2130. .2186
                     /note="lac promoter"
     gene            2207. .3243
                     /gene="bioB"
     RBS             2207. .2211
                     /gene="bioB"
     CDS             2218. .3243
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAA81773.1"
                     /db_xref="GI:1073248"
                     /translation="MTMITNSHRMRDSVIDIRGLDRVQRARVQAKEEAPKRWSVDDIV
                     ALFELPFSDLMHRAQSVHRENFDPNGVQVSTLLSIKTGGCSEDCGYCPQAARYHTDVE
                     KQDLHAAGRGTRSRVLPRRMRQPFCMAAWQPKQRDLEPVLAMIREVKAMGLETCATLG
                     MLKDGQAEQLKEAGLDYYNHNLDTAPEYYGEVITTRTYQDRLDTLDRVREQDINVCCG
                     GIIGMGESRVQRAGLLAQLANMERPPESVPDQLLTQVEGTPMYGMDELDPFEFVRTIA
                     AARITMPQSFVRLSVGRQSMHEGIQALCFLAGANSIFYGEKLLTTGNPEAEADRKLFD
                     KLGIHPL"
     misc_feature    2245. .3315
                     /note="isolated from Methylobacillus flagellatum"
     misc_feature    join(3321. .3748,1. .2232)
                     /note="derived from pUC18"
BASE COUNT      895 a    980 c    988 g    885 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 12;  Length 3748;
  Best Local Similarity   90.5%;  Pred. No. 8.8e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcc 21
           ||||||| || ||||||||||
Db    2400 GCATCGCGAGAATTTCGATCC 2420


RESULT  36
AE002396
LOCUS       AE002396    12107 bp    DNA             BCT       25-MAY-2000
DEFINITION  Neisseria meningitidis serogroup B strain MC58 section 38 of 206 of
            the complete genome.
```

```
ACCESSION   AE002396 AE002098
VERSION     AE002396.2  GI:7413429
KEYWORDS    .
SOURCE      Neisseria meningitidis MC58.
  ORGANISM  Neisseria meningitidis MC58
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 12107)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Complete genome sequence of Neisseria meningitidis serogroup B
            strain MC58
  JOURNAL   Science 287 (5459), 1809-1815 (2000)
  MEDLINE   20175755
   PUBMED   10710307
REFERENCE   2  (bases 1 to 12107)
  AUTHORS   Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
            Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
            Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
            Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
            Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
            Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
            Qin,H., Vamathevan,J., Gill,J., Scarlato,V., Masignani,V.,
            Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
            Rappuoli,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On Apr 4, 2000 this sequence version replaced gi:7225622.
FEATURES             Location/Qualifiers
     source          1. .12107
                     /organism="Neisseria meningitidis MC58"
                     /strain="MC58"
                     /db_xref="taxon:122586"
                     /note="serogroup: B"
     gene            168. .740
                     /gene="NMB0400"
                     /note="This region contains a match to at least one other
                     gene that is not full length,and is not the result of a
                     sequencing artifact; transposase, truncated; identified by
                     Glimmer2; putative;transposase, truncated"
     gene            complement(1774. .5379)
                     /gene="NMB0401"
     CDS             complement(1774. .5379)
                     /gene="NMB0401"
                     /note="similar to PID:1262198 percent identity: 70.26;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="proline dehydrogenase"
                     /protein_id="AAF40840.1"
                     /db_xref="GI:7225623"
                     /translation="MFHFAFPAQTALRQAITDAYRRNEIEAVQDMLQRAQMSDEERNA
                     ASELARRLVTQVRAGRTKAGGVDALMHEFSLSSEEGIALMCLAEALLRIPDNATRDRL
                     IADKISDGNWKSHLNNSPSLFVNAAAWGLLITGKLTATNDKQMSSALSRLISKGGAPL
                     IRQGVNYAMRLLGKQFVTGQTIEEALQNGKEREKMGYRFSFDMLGEAAYTQADADRYY
                     RDYVEAIHAIGKDAAGQGVYEGNGISVKLSAIHPRYSRTQHGRVMGELLPRLKELFLI
                     GKKYDIGINIDAEEANRLELSLDLMEALVSDPDLAGYKGIGFVVQAYQKRCPFVIDYL
                     IDLARRNNQKLMIRLVKGAYWDSEIKWAQVDGLNGYPTYTRKVHTDISYLACARKLLS
                     AQDAVFPQFATHNAYTLGAIYQMGKGKDFEHQCLHGMGETLYDQVVGPQNLGRRVRVY
                     APVGTHETLLAYLVRRLLENGANSSFVNQIVDENISIDTLIRSPFDTIAEQGIHLHNA
                     LPLPRDLYGKCRLNSQGVDLSNENVLQQLQEQMNKAAAQDFHAASIVNGKARDVGEAQ
                     PIKNPADHDDIVGTVSFADAALAQEAVGAAVAAFPEWSATPAAERAACLRRFADLLEQ
                     HTPALMMLAVREAGKTLNNAIAEVREAVDFCRYYANEAEHTLPQDAKAVGAIVAISPW
                     NFPLAIFTGEVVSALAAGNTVIAKPAEQTSLIAGYAVSLMHEAGIPTSALQLVLGAGD
                     VGAALTNDARIGGVIFTGSTEVARLINKALAKRGDNPVLIAETGGQNAMIVDSTALAE
                     QVCADVLNSAFDSAGQRCSALRILCVQEDVADRMLDMIKGAMDELVVGKPIQLTTDVG
                     PVIDAEAQQNLLNHINKMKGVAKSYHEVKTAADVDSKKSTFVRPILFELNNLNELQRE
                     VFGPVLHVVRYRADELDNVIDQINSKGYALTHGVHSRIEGTVRHIRSRIEAGNVYVNR
                     NIVGAVVGVQPFGGHGLSGTGPKAGGSFYLQKLTRAGEWVAPTLSQIGQADEAALKRL
                     EALVHKLPFNAEEKKAAAAALGHARIRTLRRAETVLTGPTGERNSISWHAPKRVWIHG
                     GSTVQAFAALTELAASGIQAVVEPDSPLASYTADLEGLLLVNGKPETAGISHVAALSP
                     LDSARKQELAAHDGALIRILPSENGLDILQVFEEISCSVNTTAAGGNASLMAVAD"
     gene            complement(5628. .7154)
                     /gene="NMB0402"
     CDS             complement(5628. .7154)
                     /gene="NMB0402"
                     /note="similar to GB:L01146 GB:L01148 GB:L01136 GB:L01140
                     GB:L01144 percent identity: 82.00; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="sodium/proline symporter"
                     /protein_id="AAF40841.1"
                     /db_xref="GI:7225624"
                     /translation="MNPMYITFAIYLVAVLLIGLAAYFSTRNFDDYILGGRSLGPFVT
                     AMSAGASDMSGWLLMGLPGAIYLSGLNEAWIAIGLLVGAYFNWLLVAGRLRVHTEYAN
                     NALTLPDYFFHRFGAGGHLMKVVSALIILFFFTIYCASGIVAGATLFQSLFEGMTYNQ
                     AMWLGAGATIAYTFLGGFLAVSWTDTLQASLMIFALILTPVMVYLGLGGAEQMSAAIQ
                     SVAAGTGKEYGSLFAGTTVIGIISTAAWGLGYFGQPHILARFMAAESAKSLVSARRIG
                     MTWMALCLAGAVAVGYFGIAYFGANPDKVSSMSGNHERIFIALSTLLFNPWIAGIILS
                     AILAAVMSTLSCQLLVCSSAITEDFYKGFLRKNAQQSELVWVGRLMVLAIAVISILIA
                     SDPNSKVLGLVSYAWAGFGAAFGPIVILSVLWKRITAYGALSGMVAGALTVVVWAEWV
                     KKPAQAAGESGLLTMYEIVPGFIVCLIVAVLVSLFNKEPSRETQERFEKADADYRATR
                     "
     gene            complement(7490. .7582)
                     /gene="NMB0403"
     CDS             complement(7490. .7582)
                     /gene="NMB0403"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF40842.1"
                     /db_xref="GI:7225625"
                     /translation="MPSENGRRIIGKTYRASFAVAKLHLKTPAA"
     gene            7637. .7957
                     /gene="NMB0404"
     CDS             7637. .7957
                     /gene="NMB0404"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF40843.1"
                     /db_xref="GI:7225626"
                     /translation="MFGKQLFEEVGSKISETIANSPAKDVEKNIKAMLGGAFNRMDLV
                     TREEFDIQQQVLIKTRTKLAALEARLEKLEAAQNPERAALEAAEAAAEEAVAEIRQQT
                     EAGE"
     gene            7971. .9467
                     /gene="NMB0405"
     CDS             7971. .9467
                     /gene="NMB0405"
                     /note="similar to GB:L42023 SP:P45049 PID:1006428
                     PID:1221236 PID:1205364 percent identity: 70.42;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="competence protein ComM"
                     /protein_id="AAF40844.1"
                     /db_xref="GI:7225627"
                     /translation="MSLALVYSRALSGMNAPLVEVEAHLANGLPHFNIVGLPDMEVKE
                     SRDRVRAAIIQSGFEFPAKKITVNLAPADLPKESGRFDLPIAIGILAASGQVAPEKLE
                     EYEFAGELALSGLLRPVRGALAMAWQGMQAKRAFVLPEENAGQAAVMRGITVYGARSL
                     GEVAAHLNGIEPLAQTECQVPQMPFEHGGQPDLCDVKGQHTARLALEIAAAGGHSLLM
                     MGPPGTGKSMLSQRLPGILPPLTEDELVEVWALRSLLPNHQQQLDSNRPFRSPHHSAS
                     AAAMVGGGSDPRPGEISLAHHGVLFLDELPEFDRKVLEVLREPLENGEIHISRAARQA
```

```
                     VYPAKFQLVAAMNPCPCGYLGHPVKPCRCTPESVARYRSKISGPLLDRIDLTIEVPSL
                     SAAELMQQEAGESSASVLERVIAARDKQYARQGKVNAALSVSELDTSARIQKEAQEAL
                     GGLLEKLSLSARSFHRIMRVARTLADLAGDEEVGRSHVMKAIGFRRAL"
     gene            9567. .10436
                     /gene="NMB0406"
     CDS             9567. .10436
                     /gene="NMB0406"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF40845.1"
                     /db_xref="GI:7225628"
                     /translation="MFMNKFSQSGKGLSGFFFGLILATVIIAGILFYLNQSGQNAFKI
                     PASSKQPAETEILKPKNQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAADKQPVAD
                     KADEVEEKAGEPEREEPDGQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKET
                     EKKASKEEKKAAKEKVAPKPTPEQILNSGSIEKARSAAAKEVQKMKTSDKAEATHYLQ
                     MGAYADRQSAEGQRAKLAILGISSKVVGYQAGHKTLYRVQSGNMSADAVKKMQDELKK
                     HEVASLIRSIESK"
     gene            10439. .11083
                     /gene="NMB0407"
     CDS             10439. .11083
                     /gene="NMB0407"
                     /note="similar to SP:P95460 PID:1814074 percent identity:
                     56.84; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="thiol:disulfide interchange protein DsbA"
                     /protein_id="AAF40846.1"
                     /db_xref="GI:7225629"
                     /translation="MKLKHLLPLLLSAVLSAQAYALTEGEDYLVLDKPIPQEQSGKIE
                     VLEFFGYFCVHCHHFDPLLLKLGKALPSDAYLRTEHVVWQPEMLGLARMAAAVNLSGL
                     KYQANPAVFKAVYEQKIRLENRSVAGKWALSQKGFDGKKLMRAYDSPEAAAAALKMQK
                     LTEQYRIDSTPTVIVGGKYRVIFNNGFDGGVHTIKELVAKVREERKRQTPAVQK"
     gene            11116. .11937
                     /gene="NMB0408"
     CDS             11116. .11937
                     /gene="NMB0408"
                     /note="similar to PID:882579 GB:U00096 PID:1789437 percent
                     identity: 67.54; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="bacitracin resistance protein"
                     /protein_id="AAF40847.1"
                     /db_xref="GI:7225630"
                     /translation="MDFLIVLKALMMGLVEGFTEFLPISSTGHLIVFGNLIGFHSNHK
                     VFEIAIQLGAVLAVVFEYRQRFSNVLHGLGKDRKANRFVLNLAIAFIPAAVMGLLFGK
                     QIKEYLFNPLSVAVMLVLGGFFILWVEKRQSRAEPKIADVDALRPIDALMIGVAQVFA
                     LVPGTSRSGSTIMGGMLWGIERKTATEFSFFLAVPMMVAATAYDVLKHYRFFTLHDVG

  Query Match             71.2%;  Score 17.8;  DB 1;  Length 12107;
  Best Local Similarity   90.5%;  Pred. No. 8.6e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcct 22
           ||| |||| ||||||||||||
Db   10601 CATTGCCATCATTTCGATCCT 10621


RESULT  37
AC018306/c
LOCUS       AC018306    14645 bp    DNA             HTG       09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC018306
VERSION     AC018306.1  GI:6552885
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 14645)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10214280 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .14645
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     4097 a    3678 c    3403 g    3467 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 2;  Length 14645;
  Best Local Similarity   90.5%;  Pred. No. 8.6e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcc 21
           |||||||||||||  ||||||
Db   10697 GCATCGCCAGCATCGCGATCC 10677


RESULT  38
AP003375/c
LOCUS       AP003375   106462 bp    DNA             HTG       27-MAR-2001
DEFINITION  Oryza sativa chromosome 1 clone OJ1414_E05, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AP003375
VERSION     AP003375.1  GI:13365592
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1414_E05.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (sites)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
            clone:OJ1414_E05
  JOURNAL   Published Only in DataBase (2001) In press
REFERENCE   2  (bases 1 to 106462)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .106462
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="1"
```

```
                     /clone="OJ1414_E05"
BASE COUNT     28948 a   24036 c   23776 g   29652 t      50 others
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 2;  Length 106462;
  Best Local Similarity   90.5%;  Pred. No. 8.3e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcc 21
           ||||||||| |||||| ||||
Db   14284 GCATCGCCACCATTTCCATCC 14264


RESULT  39
TBBCHR1A_03
WPCOMMENT
  Sequence split into 10 fragments  LOCUS TBBCHR1A Accession AL359782
     Fragment Name      Begin     End
     TBBCHR1A_00            1     110000
     TBBCHR1A_01       100001     210000
     TBBCHR1A_02       200001     310000
     TBBCHR1A_03       300001     410000
     TBBCHR1A_04       400001     510000
     TBBCHR1A_05       500001     610000
     TBBCHR1A_06       600001     710000
     TBBCHR1A_07       700001     810000
     TBBCHR1A_08       800001     910000
     TBBCHR1A_09       900001     941631
Continuation (4 of 10) of TBBCHR1A from base 300001 (AL359782 Trypanosoma brucei chromos


  Query Match             71.2%;  Score 17.8;  DB 2;  Length 110000;
  Best Local Similarity   90.5%;  Pred. No. 8.3e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcc 21
           ||||||||||| ||| ||||
Db   41181 GCATCGCCAGCAATTCCATCC 41201


RESULT  40
AC093192
LOCUS       AC093192   166857 bp    DNA             INV       16-AUG-2001
DEFINITION  Drosophila melanogaster  BAC RP98-48J18 (Roswell Park Cancer
            Institute Drosophila BAC Library) complete sequence.
ACCESSION   AC093192
VERSION     AC093192.1  GI:15193302
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 166857)
  AUTHORS   Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,
            Gocayne,J.D., Tabor,P., Williamson,A., Homsi,F.H.,
            Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
            Ayele,M.A., Scott,G.S., Worley,K.W., Amamatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
            Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
            Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
            Draper,H., Emery-Cohen,A., Ferriera,S., Garg,N.D.S., Houck,J.,
            Hostin,D., Howland,T.J., Hume,J., Ibeqwam,C., Jalali,M., Kovar,C.,
            Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
            Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,
            Scheeler,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,
            Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,
            Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166857)
  AUTHORS   Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
```

```
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
            html.

            QUALSTAT-REPORT.
FEATURES              Location/Qualifiers
     source           1. .166857
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="RP98-48J18"
     misc_feature     complement(22580. .166857)
                      /note="Overlaps bases 172173. .27896 of clone AC010038"
                      /function="clone overlap"
BASE COUNT    42310 a  39417 c  39558 g  45572 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 3;  Length 166857;
  Best Local Similarity   90.5%;  Pred. No. 8.3e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcc 21
          |||||||||||||  ||||||
Db 135217 GCATCGCCAGCATCGCGATCC 135237


RESULT  41
AC007809
LOCUS       AC007809   183439 bp    DNA             INV       24-FEB-2001
DEFINITION  Drosophila melanogaster, chromosome 3R, region 88C-88C, BAC clone
            BACR45M03, complete sequence.
ACCESSION   AC007809
VERSION     AC007809.8  GI:13122709
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 183439)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 3R, region 88C-88C
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 183439)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Feb 24, 2001 this sequence version replaced gi:6623906.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES              Location/Qualifiers
     source           1. .183439
                      /organism="Drosophila melanogaster"
                      /strain="y; cn bw sp"
                      /db_xref="taxon:7227"
                      /chromosome="3R"
                      /map="88C-88C"
                      /clone="BACR45M03 (D718)"
                      /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                      Drosophila melanogaster BAC library, partial EcoRI in
                      pBACe3.6)"
BASE COUNT    53300 a  38849 c  38854 g  52436 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 3;  Length 183439;
  Best Local Similarity   90.5%;  Pred. No. 8.3e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 catcgccagcatttcgatcct 22
          | ||||||||||||||||| |
Db  66705 CCTCGCCAGCATTTCGATCAT 66725


RESULT  42
NMA7Z2491/c
LOCUS       NMA7Z2491  195767 bp    DNA             BCT       04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 7/7.
ACCESSION   AL162758 AL157959
VERSION     AL162758.2  GI:7380672
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 195767)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 195767)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES              Location/Qualifiers
     source           1. .195767
                      /organism="Neisseria meningitidis Z2491"
                      /strain="Z2491"
                      /db_xref="taxon:122587"
                      /note="serogroup: A"
     gene             80. .952
                      /gene="NMA2045"
```

```
CDS             80. .952
                /gene="NMA2045"
                /note="NMA2045, possible oxidoreductase, len: 290aa;
                similar to many eg. SW:P20692 (TYRA_BACSU) prephenate
                dehydrogenase from Bacillus subtilis (372 aa) fasta
                scores; E(): 2.8e-30, 39.4% identity in 284 aa overlap."
                /codon_start=1
                /transl_table=11
                /product="putative oxidoreductase"
                /protein_id="CAB85263.1"
                /db_xref="GI:7380673"
                /db_xref="SPTREMBL:Q9JT12"
                /translation="MPILNHIALIGVGLIGGSFVLDLKRQGLVRTVTGIDTDRDNLER
                ALERGVIDRASVAIDADSIGGADLVLIATPVATVPAILTALRPVLPEHTWISDVGSTK
                SSVIEAFRRCLPDRLHRCIAAHPIAGSDRSGAQAAQFGLFRHRKLIITPHGGEHSDGI
                ALIENLWHAVGAEIYTMDAQRHDAVFAAVSHMPHLTAFAYVHQILDHPDGQEYLKFAA
                TGFRDFTRIASGHPAVWADICLANKDSLLQLVQGLGKQLDVLADILTADDREALYRYF
                EEAKTTRDRWLDGN"
misc_feature    complement(547. .556)
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
misc_feature    960. .969
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
gene            complement(1009. .2211)
                /gene="NMA2046"
CDS             complement(1009. .2211)
                /gene="NMA2046"
                /note="NMA2046, hypothetical integral membrane protein,
                len: 400aa; similar to TR:P95556 (EMBL:Z73914)
                hypothetical protein from Pseudomonas stutzeri (396 aa)
                fasta scores; E(): 0, 35.0% identity in 391 aa overlap.
                Contains possible membrane-spanning hydrophobic regions."
                /codon_start=1
                /transl_table=11
                /product="hypothetical integral membrane protein"
                /protein_id="CAB85264.1"
                /db_xref="GI:7380674"
                /db_xref="SPTREMBL:Q9JT11"
                /translation="MIPFCGKYINDKQGTHMKFTKHPVWAMAFRPFYSLAALYGALSV
                LLWGFGYTGTHELSGFYWHAHEMIWGYAGLVVIAFLLTAVATWTGQPPTRGGVLVGLT
                IFWLAARIAAFIPGWGASASGILGTLFFWYGAVCMALPVIRSQNQRNYVAVFALFVLG
                GTHAAFHVQLHNGNLGGLLSGLQSGLVMVSGFIGLIGTRIISFFTSKRLNVPQIPSPK
                WVAQASLWLPMLTAMLMAHGVMPWLSAAFAFAAGVIFTVQVYRWWYKPVLKEPMLWIL
                FAGYLFTGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLTLGMMARTALGHTGNPIYP
                PPKAVPVAFWLMMAATAVRMVAVFSSGTAYTHSIRTSSVLFALALLVYAWKYIPWLIR
                PRSDGRPG"
gene            complement(2208. .2576)
                /gene="NMA2047"
CDS             complement(2208. .2576)
                /gene="NMA2047"
                /note="NMA2047, hypothetical protein, len: 122aa; weakly
                similar to SW:P71932 (YQ03_MYCTU) hypothetical protein
                from Mycobacterium tuberculosis (161 aa) fasta scores;
                E(): 0.017, 32.5% identity in 117 aa overlap."
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="CAB85265.1"
                /db_xref="GI:7380675"
                /db_xref="SPTREMBL:Q9JT10"
                /translation="MKPLKRHPALIGLSRDHHHSLSLCVRLLRTPEERHRDELEPHFS
                ELETHFREEEAKFAPIWQNVAPELKQRFEKDHARLRQMMASPEYGNAAWNTAFATTLR
                DHARFEERELFPAVEPFLPA"
misc_feature    complement(2622. .2631)
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
gene            2750. .3193
                /gene="NMA2048"
CDS             2750. .3193
                /gene="NMA2048"
                /note="NMA2048, conserved hypothetical protein, len:
                147aa; similar to many eg. SW:P21498 (YJEB_ECOLI)
                hypothetical protein from Escherichia coli (141 aa) fasta
                scores; E(): 5.9e-23, 47.9% identity in 140 aa overlap."
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="CAB85266.1"
                /db_xref="GI:7380676"
                /db_xref="SPTREMBL:Q9JT09"
                /translation="MYLTQHTDYGLRVLIYTAINDDALVNISTIAVTYGISKSHLMKV
                VTALVKGGFLHSVRGKGGGLRLASPPERINIGAVVRHLEPMQLVECMGENNECLITPS
                CRLTGILGGAMKSFFTYLDGFTLQDLLNKPTYDLLYEPRIPIAVQ"
stem_loop       3275. .3314
                /note="Stem loop containing DNA uptake sequences: gcaat
                gccgtctgaa cagccgatgc ttcagacggc atcgg"
misc_feature    3280. .3289
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
misc_feature    complement(3300. .3309)
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
gene            complement(3320. .3940)
                /gene="NMA2049"
CDS             complement(3320. .3940)
                /gene="NMA2049"
                /note="NMA2049, conserved hypothetical protein, len:
                206aa; similar to many eg. SW:P45847 (YCIO_ECOLI)
                hypothetical protein from Escherichia coli (206 aa) fasta
                scores; E(): 0, 62.9% identity in 205 aa overlap. Contains
                Pfam match to entry PF01300 Sua5_yciO_yrdC, SUA5/yciO/yrdC
                family and Prosite match to PS01147 SUA5/yciO/yrdC family
                signature."
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="CAB85267.1"
                /db_xref="GI:7380677"
                /db_xref="SPTREMBL:Q9JT08"
                /translation="MAQFFAIHPDNPQERLIKQAVEIVNKGGVVVYPTDSCYALGCKL
                GDKAAMERILSIRKIDLKHHLTLMCADLSELGTYAKIDNVQFRQLKAATPGPYTFILQ
                ATKDVPARTLHPKRKTIGLRIPDNAIAQALLGELGEPLLSCTLMLPEDGEPLTDPYEI
                RECLEHAVDLVIDGGWCGTEPTTVVDMTDGTELVRQGCGDTAVFGL"
misc_feature    complement(3350. .3877)
                /gene="NMA2049"
                /note="Pfam match to entry PF01300 Sua5_yciO_yrdC,
                SUA5/yciO/yrdC family, score 192.00, E-value 9.6e-54"
misc_feature    complement(3818. .3856)
                /gene="NMA2049"
                /note="PS01147 SUA5/yciO/yrdC family signature."
repeat_unit     4027. .4046
                /note=">= 90% match to ATTCCCNNNNNNNNGGGAAT"
                /label=dRS3
repeat_unit     4047. .4102
                /note="85% identity to consensus
                CCAGTCCGTTCAGTTTCAGTCATTTCCGATAAATTCCTGTTGCTTTTCATTTCTAG"
                /label=RS17
misc_feature    complement(4262. .4271)
                /note="Core DNA uptake sequence: gccgtctgaa"
                /label=DUS
gene            complement(4298. .5497)
                /gene="ackA2"
CDS             complement(4298. .5497)
                /gene="ackA2"
                /EC_number="2.7.2.1"
                /note="NMA2050, ackA2, acetate kinase, len: 399aa; similar
                to many eg. SW:P15046 (ACKA_ECOLI) ackA, acetate kinase
                from Escherichia coli (400 aa) fasta scores; E(): 0, 49.5%
                identity in 400 aa overlap. Also similar to NMA1718 from
                Neisseria meningitidis fasta scores; E(): 0, 63.3%
                identity in 398 aa overlap. Contains Pfam match to entry
                PF00871 Acetate_kinase, Acetokinase family; Prosite match
                to PS01075 Acetate and butyrate kinases family signature 1
                and Prosite match to PS01076 Acetate and butyrate kinases
                family signature 2."
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="acetate kinase"
                     /protein_id="CAB85268.1"
                     /db_xref="GI:7380678"
                     /db_xref="SPTREMBL:Q9JT07"
                     /translation="MSDQLILVLNCGSSSLKGAVIDRKSGSVVLSCLGERLTTPEAVI
                     TFSKDGNKRQVPLSGRNCHAGAVGMLLNELEKHELHDRIQAVGHRIAHGGEKYSESVL
                     IDQAVMDELNACIPLAPLHNPANISGILAAQEHFPGLPNVGVMDTSFHQTMPERAYTY
                     AVPRELRKKYAFRRYGFHGTSMRYVAPEAACILGKPLEDIRMIIAHLGNGASITAIKN
                     GKSVDTSMGFTPIEGLVMGTRCGDIDPGVYSYLTSHAGLDVAQVDEMLNKKSGLLGIS
                     ELSNDCRTLEIAADEGHEGARLALEVMTYRLAKYIASMAVGCGGVDALVFTGGIGENS
                     RNIRAKTVSYLDFLGLHIDTKANMEKRYGNSGIISPTDSSPAVLVVPTNEELMIACDT
                     AELVGIL"
     misc_feature    complement(4328. .5485)

  Query Match             71.2%;  Score 17.8;  DB 1;  Length 195767;
  Best Local Similarity   90.5%;  Pred. No. 8.3e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcct 22
           ||| |||| ||||||||||||
Db   38896 CATTGCCATCATTTCGATCCT 38876


RESULT  43
AE003705
LOCUS       AE003705   232744 bp    DNA             INV       05-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386035 section 30
            of 105, complete sequence.
ACCESSION   AE003705 AE002708
VERSION     AE003705.1  GI:7299886
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 232744)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
    TITLE   The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 232744)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
    TITLE   Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES             Location/Qualifiers
     source          1. .232744
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
     mRNA            complement(join(29274. .29439,29524. .30038,30107. .>30980))
                     /gene="CG8464"
                     /product="CT24781"
                     /db_xref="FLYBASE:FBan0008464"
                     /db_xref="FLYBASE:FBgn0038233"
     gene            complement(<29274. .>30980)
                     /gene="CG8464"
                     /map="88C4-88C6"
                     /db_xref="FLYBASE:FBan0008464"
                     /db_xref="FLYBASE:FBgn0038233"
     CDS             complement(join(29274. .29439,29524. .30038,30107. .30694))
                     /gene="CG8464"
                     /note="CG8464 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0008464"
                     /db_xref="FLYBASE:FBgn0038233"
                     /protein_id="AAF55062.1"
                     /db_xref="GI:7299887"
                     /translation="MALRGSHRLEVIFKRCIASPVLHSHAANRRSSQLAIKEGDPNSN
                     GNSGQYQQNGEQKEKGWRRLVRFFVPFSLGAVVSAAIIQREDLTPTIAASKMTGRRRD
                     FNFIADVVAGCADSVVYIEIKDTRHFDYFSGQPITASNGSGFIIEQNGLILTNAHVVI
                     NKPHTMVQVRLSDGRTFPATIEDVDQTSDLATLRIQVNNLSVMRLGKSSTLRSGEWVV
                     ALGSPLALSNTVTAGVISSTQRASQELGLRNRDINYLQTDAAITFGNSGGPLVNLDGE
                     AIGVNSMKVTAGISFAIPIDYVKVFLERAAEKRKKGSAYKTGYPVKRYMGITMLTLTP
                     DILFELKSRSQNMPSNLTHGVLVWKVIVGSPAHSGGLQPGDIVTHINKKEIKNSSDVY
                     DALADNSKTLDIVILRGVKQMHVTITPEDP"
     mRNA            join(<31212. .31346,31407. .31691,31752. .>31922)
                     /gene="CG3351"
                     /product="CT11255"
                     /db_xref="FLYBASE:FBan0003351"
                     /db_xref="FLYBASE:FBgn0038234"
     gene            <31212. .>31922
                     /gene="CG3351"
                     /map="88C4-88C6"
                     /db_xref="FLYBASE:FBan0003351"
                     /db_xref="FLYBASE:FBgn0038234"
     CDS             join(31212. .31346,31407. .31691,31752. .31922)
                     /gene="CG3351"
                     /note="CG3351 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0003351"
                     /db_xref="FLYBASE:FBgn0038234"
                     /protein_id="AAF55063.1"
                     /db_xref="GI:7299888"
                     /translation="MSKAAGKLKSLKKTVERVTHTSKLKTNIPAGMAAAGPPLGPMLG
                     QRAINIAAFCKDFNAKTAEMKEGVPLPCRISVNSDRSYDLAIHHPPATFFLKQAAGIQ
                     RGTMTPGKEVAGMITLKHLYEIAAIKIQDPPNVLLTMQQMCEMLISIARTCGIKVVRE
                     IDPAAYGEFLEERKLIVEQQRRELQEKREAKMLRTG"
     mRNA            complement(join(32136. .32708,32770. .>32889))
                     /gene="CG8461"
                     /product="CT24741"
                     /db_xref="FLYBASE:FBan0008461"
                     /db_xref="FLYBASE:FBgn0038235"
     gene            complement(<32136. .>32889)
                     /gene="CG8461"
                     /map="88C4-88C6"
```

```
                /db_xref="FLYBASE:FBan0008461"
                /db_xref="FLYBASE:FBgn0038235"
     CDS        complement(join(32136. .32708,32770. .32835))
                /gene="CG8461"
                /note="CG8461 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0008461"
                /db_xref="FLYBASE:FBgn0038235"
                /protein_id="AAF55064.1"
                /db_xref="GI:7299889"
                /translation="MSSPSESSSASEHEASDDNETQNAIREDLKGMSFEEIMKLKEEL
                GAKVYKEAVLGSKSSRPQKPKAKTDLKRLNKNRPREMTTRRQVPFLGAEHRVERKKNV
                ELRDPRFDEKSGTYSAETFKKNYQFVSQIRAKEVGQLKKKLDEVEDEAEKHHIKDTMQ
                RLINKNVEDKKWHTKQKQLKKERSAIQKKHDLGQQPHYLTKSTFLIDSEIKF"
     mRNA       complement(join(<34919. .35540,35598. .35757,35894. .36047,
                36334. .36679,36886. .37166,37233. .37324,37381. .37624,
                37694. .37981,38058. .38138,42442. .42707,42765. .43010,
                44497. .44761,48465. .>49118))
                /gene="CG8279"
                /product="CT24224"
                /db_xref="FLYBASE:FBan0008279"
                /db_xref="FLYBASE:FBgn0038237"
     gene       complement(<34919. .>49118)
                /gene="CG8279"
                /map="88C5-88C6"
                /db_xref="FLYBASE:FBan0008279"
                /db_xref="FLYBASE:FBgn0038237"
     CDS        complement(join(34919. .35540,35598. .35757,35894. .36047,
                36334. .36679,36886. .37166,37233. .37324,37381. .37624,
                37694. .37981,38058. .38138,42442. .42707,42765. .43010,
                44497. .44761,48465. .49118))
                /gene="CG8279"
                /note="CG8279 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0008279"
                /db_xref="FLYBASE:FBgn0038237"
                /protein_id="AAF55066.2"
                /db_xref="GI:10726529"
                /translation="MHGPVSRSSSSSNMTDVSSPAGGAASPVEMSTTSSSSAATTSAS
                SSKPLTNGANKTAISTAAGGVTPGAVPGPGSGAIPASSSSGNQVKLEHHHRQSNNNRP
                AVTNRSSETKLMTPTGSSSSPSQSPSQTQASIQTQTSQQDRLAKASTTASQQDVDEVA
                RLFEEKPEAFEKWLTERAPPEALSRLQEFIENRKPHKRPSVTSDLFQQWMAASPTVQQ
                KSPRSLSNSSASSLPECRRHLMDLDEGELFMELIRDVANELDIDVLCHKILVNVGLLT
                HADRGSLFLAKGTPTNKYLVAKLFDVTQKTALKDAVTRASAEEIIIPFGIGIAGMVAQ
                TKQMINIKEAYKDARFNCEIDLKTGYKTNAILCMPICNYEGDIIGVAQIINKTNGCME
                FDEHDVEIFRRYLTFCGIGIQNAQLFEMSVQEYRRNQILLNLARSIFEEQNNLECLVT
                KIMTEARELLKCERCSVFLVDLDCCEASHLEKIIEKPNQPATRAIKSADSFEEKKMRN
                RFTVLFELGGEYQAANVSRPSVSELSSSTLAQIAQFVATTGQTVNICDVIEWVRDHNQ
                IRAEDEIDSTQAILCMPIMNAQKKVIGVAQLINKANGVPFTDSDASIFEAFAIFCGLG
                IHNTQMYENACKLMAKQKVALECLSYHATASQDQTEKLTQDVIAEAESYNLYSFTFTD
                FELVDDDTCRAVLRMFMQCNLVSQFQIPYDVLCRWVLSVRKNYRPVKYHNWRHALNVA
                QTMFAMLKTGKMERFMTDLEILGLLVACLCHDLDHRGTNNAFQTKTESPLAILYTTST
                MEHHHFDHFRLAAFLDRLAAFRRALRDLARELLDLVEELVVDPTLIFLVIEYFRLLRR
                TIFLFEPELDEEDDDVVDSVVDSVVVLSVVLVLDSVLVLLSAALTWATYKTIERTKRT
                RDNFMLDALSPEDYRSVMKTVESAILSTDLAMYFKKRNAFLELVENGEFDWQGEEKKD
                LLCGMMMTACDVSAIAKPWEVQHKVAKLVADEFFDQGDLEKLQLNTQPVAMMDRERKD
                ELPKMQVGFIDVICLPLYRVLCDTFPWITPLYEGTLENRRNWQDLAEKVEMGLTWIDH
                DTIDKPVEEFAACADEEIKDIEFTVTTLNCNQQSQHGSEDSHTPEHQRSGSRLSMKKT
                GALGKAVRSKLSKTLYNSMDGSKPKTSLKLLESHVSEDMDDKSPTSPSQPQASGSMGR
                MSASSSTSSAGGQMVDKSKKRSKLCALL"
     mRNA       join(40359. .40547,40626. .40826,40891. .41076,41138. .41210,
                41284. .41410,41461. .42034,42102. .>42262)
                /gene="Cyp313a1"
                /product="CT11271"
                /db_xref="FLYBASE:FBan0003360"
                /db_xref="FLYBASE:FBgn0038236"
     gene       <40359. .>42262
                /gene="Cyp313a1"
                /note="CG3360"
                /map="88C5-88C6"
                /db_xref="FLYBASE:FBan0003360"
                /db_xref="FLYBASE:FBgn0038236"
     CDS        join(40391. .40547,40626. .40826,40891. .41076,41138. .41210,
                41284. .41410,41461. .42034,42102. .42262)
                /gene="Cyp313a1"
                /note="Cyp313a1 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0003360"
                /db_xref="FLYBASE:FBgn0038236"
                /protein_id="AAF55065.2"
                /db_xref="GI:10726530"

  Query Match             71.2%;  Score 17.8;  DB 3;  Length 232744;
  Best Local Similarity   90.5%;  Pred. No. 8.2e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcct 22
           | ||||||||||||||||||| |
Db 106436 CCTCGCCAGCATTTCGATCAT 106456


RESULT  44
AC017740/c
LOCUS       AC017740   237119 bp    DNA             HTG       10-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017740
VERSION     AC017740.1  GI:6554259
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 237119)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211941 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .237119
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    67689 a  49909 c  50172 g  69349 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 2;  Length 237119;
  Best Local Similarity   90.5%;  Pred. No. 8.2e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcct 22
           | ||||||||||||||||||| |
Db 103683 CCTCGCCAGCATTTCGATCAT 103663


RESULT  45
AC005557
LOCUS       AC005557   257867 bp    DNA             INV       14-NOV-1998
DEFINITION  Drosophila melanogaster, chromosome 3L, region 62A10-62B5, P1
            clones DS02777, DS03222, DS02345, and DS04808, complete sequence.
ACCESSION   AC005557 AC004346 AC004344 AC004354 AC004332
VERSION     AC005557.1  GI:3492862
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 257867)
```

```
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Punch,E., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Twomey,B., Wan,K.H., Weinburg,T., Zhang,R.,
            Zieran,L.L. and Rubin,G.M.
  TITLE     Sequencing of Drosophila chromosome 3L, region 62A10-62B5
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 257867)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
            Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
            Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
            Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
            Yee,A., Zhang,R., Zieran,L.L. and Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-1998) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
            P1 library locations: 29-89, 34-54, 25-41.
FEATURES             Location/Qualifiers
     source          1. .257867
                     /organism="Drosophila melanogaster"
                     /strain="y2; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="3L"
                     /map="62A10-62B5"
                     /clone_lib="P1 library, partial Sau3A in pAd10sacBII"
                     /clone="P1s DS02777 (D245), DS03222 (D243), DS02345
                     (D257), and DS04808 (D228)"
                     /note="DS02777 (D245) extends from bp 1 to bp 78,221.
                     DS03222 (D243) is a bridge P1 extending from bp 78,210 to
                     P1 end at bp 142,019.  DS02345 (D257) is a bridge P1
                     extending  from bp 141,146 to bp  167,538.  DS04808 (D228)
                     extends from bp 170,514 to bp 257,867."
BASE COUNT    66126 a  61150 c  61220 g  69371 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 3;  Length 257867;
  Best Local Similarity   90.5%;  Pred. No. 8.2e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcc 21
           ||||||||||||  ||||||
Db   49770 GCATCGCCAGCATCGCGATCC 49790



Search completed: December 26, 2001, 10:58:04
Job time: 2046 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:19:02 ; Search time 366.42 Seconds
(without alignments)
58.493 Million cell updates/sec

Title: US-09-396-196F-7
Perfect score: 25
Sequence: 1 gcatcgccagcatttcgatcctcgt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	25	100.0	839	22	AAI62941	Human genomic DNA
2	25	100.0	1041	20	AAX01303	E. coli biotin syn
3	25	100.0	1084	10	AAN91329	E.coli Bio B gene.
4	25	100.0	1121	7	AAN60496	Sequence encoding
5	25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
c 6	18.2	72.8	3451	18	AAV02308	Cell membrane prot
7	17.8	71.2	645	21	AAZ54339	Neisseria gonorrhe
8	17.8	71.2	645	21	AAZ54340	Neisseria meningit
9	17.8	71.2	645	21	AAZ54341	Neisseria meningit
c 10	17.8	71.2	69936	21	AAA81479	N. meningitidis pa
11	17.8	71.2	349980	21	AAF21607	Neisseria meningit
12	17.8	71.2	1437668	21	AAA81490	N. meningitidis B
c 13	17.6	70.4	20561	20	AAX13229	Enterococcus faeca
14	17.4	69.6	6386	20	AAX99364	Maize shrunken-1 (
15	16.8	67.2	61	20	AAZ19606	Complement system
16	16.8	67.2	1497	22	AAF61015	P. putida KT2440-a
c 17	16.6	66.4	609	21	AAC34251	Arabidopsis thalia
c 18	16.6	66.4	671	21	AAF13337	Aspergillus oryzae
19	16.2	64.8	455	22	AAH82438	Human ovarian tumo
c 20	16.2	64.8	501	18	AAT79955	Fragment of cb11A
c 21	16.2	64.8	543	21	AAC53704	Arabidopsis thalia
c 22	16.2	64.8	675	21	AAF12787	Aspergillus oryzae
c 23	16.2	64.8	1451	21	AAC54459	Arabidopsis thalia
c 24	16.2	64.8	1452	21	AAC54360	Arabidopsis thalia
c 25	16.2	64.8	1546	21	AAC39263	Arabidopsis thalia
c 26	16.2	64.8	1549	21	AAC54458	Sorangium cellulos
27	16.2	64.8	68750	21	AAZ55887	Oligonucleotide us
28	16	64.0	58	22	AAH42540	Pinus radiata 1,3-
c 29	16	64.0	205	21	AAA67164	Pinus radiata 1,3-
c 30	16	64.0	297	21	AAA67162	Pinus radiata 1,3-
c 31	16	64.0	393	21	AAA67165	Pinus radiata 1,3-
c 32	16	64.0	394	21	AAA67166	Pinus radiata endo
c 33	16	64.0	399	21	AAA67357	Pinus radiata 1,3-
c 34	16	64.0	419	21	AAA67169	Pinus radiata endo
c 35	16	64.0	469	21	AAA67358	Pinus radiata 1,3-
c 36	16	64.0	498	21	AAA67167	C glutamicum codin
37	16	64.0	699	22	AAH66006	Corynebacterium gl
38	16	64.0	822	22	AAF67973	Aspergillus oryzae
39	16	64.0	868	21	AAF12760	Human fascin DNA f
40	16	64.0	943	22	AAH48556	Arabidopsis thalia
c 41	16	64.0	1135	21	AAC36656	Human fascin DNA f
42	16	64.0	1240	22	AAH48555	Mouse nuclear rece
43	16	64.0	1280	20	AAX80215	Mouse CAR receptor
44	16	64.0	1361	20	AAX24003	Human fascin DNA f
45	16	64.0	1449	22	AAH48554	Human fascin DNA f
46	16	64.0	1619	22	AAH48553	Human fascin DNA f
47	16	64.0	1934	22	AAH48552	Phaffia carotenoid
48	16	64.0	2470	18	AAT72944	Phaffia carotenoid
49	16	64.0	2546	18	AAT72941	Human fascin DNA f
50	16	64.0	3069	22	AAH48551	Phaffia carotenoid
51	16	64.0	3550	18	AAT72946	Human fascin DNA f
52	16	64.0	4100	22	AAH48581	Orf virus genomic
c 53	16	64.0	4372	18	AAT91360	Plasmid pHPB12 enc
c 54	16	64.0	7475	12	AAQ15312	Human IL-1ra BAC c
55	16	64.0	9158	20	AAX03047	Mycobacterium smeg
56	16	64.0	9960	19	AAV58939	Mycobacterium tube
57	16	64.0	10095	19	AAV58938	Human fascin DNA f
58	16	64.0	13055	22	AAH48621	Human fascin DNA f
59	16	64.0	16951	22	AAH48620	Human fascin DNA f
60	16	64.0	16951	22	AAH48622	Streptococcus pneu
61	16	64.0	18475	19	AAV52171	Soybean 240017 reg
c 62	16	64.0	335913	22	AAI61371	Soybean 240017 reg
c 63	16	64.0	335913	22	AAI61372	C glutamicum codin
64	16	64.0	349980	22	AAH68527	BAC containing rep
65	15.8	63.2	151826	21	AAF22291	Drosophila gustato
66	15.6	62.4	381	22	AAF63773	EST clone CR1044.
c 67	15.6	62.4	447	20	AAV89685	Fusarium venenatum
68	15.6	62.4	629	21	AAF09259	cDNA encoding a hu
69	15.6	62.4	693	21	AAA75126	Human cDNA clone (
c 70	15.6	62.4	760	22	AAH06704	Human cDNA clone (
c 71	15.6	62.4	849	22	AAH08275	Mouse melanocortin
72	15.6	62.4	978	19	AAV19143	Mouse melanocortin
73	15.6	62.4	978	20	AAV63708	Human EST-derived
c 74	15.6	62.4	981	22	AAH98780	Human polynucleoti
c 75	15.6	62.4	981	22	AAI59815	Human polynucleoti
c 76	15.6	62.4	1046	22	AAI58029	C glutamicum codin
c 77	15.6	62.4	1098	22	AAH67893	Pituitary somatost
78	15.6	62.4	1110	13	AAQ29155	Human somatostatin
79	15.6	62.4	1130	17	AAT09250	Human somatostatin
80	15.6	62.4	1351	14	AAQ45655	Arabidopsis thalia
c 81	15.6	62.4	1905	21	AAC39550	C glutamicum codin
c 82	15.6	62.4	2715	22	AAH66459	Human cDNA sequenc
c 83	15.6	62.4	2891	22	AAH18199	Human cDNA sequenc
c 84	15.6	62.4	3219	22	AAH18553	

c	85	15.6	62.4	3600	22	AAC66910
c	86	15.6	62.4	3614	20	AAX07425
c	87	15.6	62.4	34001	22	AAF28531
c	88	15.6	62.4	349980	22	AAH68528
c	89	15.6	62.4	349980	22	AAH68533
	90	15.4	61.6	232	20	AAX51678
c	91	15.4	61.6	312	18	AAT45452
	92	15.4	61.6	333	21	AAC04391
c	93	15.4	61.6	344	21	AAC37112
c	94	15.4	61.6	396	21	AAC01432
c	95	15.4	61.6	438	21	AAC54113
c	96	15.4	61.6	440	21	AAA78595
c	97	15.4	61.6	466	21	AAC00187
c	98	15.4	61.6	486	21	AAC00188
c	99	15.4	61.6	539	22	AAF98383
c	100	15.4	61.6	557	20	AAV87119

Human EXMAD-21 cod
Homo sapiens secre
Genomic fragment #
C glutamicum codin
C glutamicum codin
Human secreted pro
Human LST-1 (leuko
Human secreted pro
Arabidopsis thalia
Human secreted pro
Arabidopsis thalia
Plant SDF polynucl
Human secreted pro
Human secreted pro
Human cDNA clone B
EST clone BO71. H

ALIGNMENTS

RESULT 1
AAI62941
ID AAI62941 standard; DNA; 839 BP.
XX
AC AAI62941;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human genomic DNA SEQ ID NO 269.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200155449-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01346.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUL-2000; 2000US-0216880.
PR 14-JUL-2000; 2000US-0218290.
PR 14-AUG-2000; 2000US-0225447.
PR 01-SEP-2000; 2000US-0229343.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231243.
PR 25-SEP-2000; 2000US-0234997.
PR 29-SEP-2000; 2000US-0236367.
PR 13-OCT-2000; 2000US-0239937.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246528.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249265.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-476225/51.
XX
PT Novel plasma membrane associated proteins useful for diagnosing,
PT treating, preventing and/or prognosing disorders related to the
PT proteins, including cancer, immune response and neuronal disorders -
XX
PS Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AAI62752-AAI62961) and proteins
CC (AAM42347-AAM42415) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;

Query Match 100.0%; Score 25; DB 22; Length 839;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatcgccagcatttcgatcctcgt 25
||||||||||||||||||||||||||
Db 122 gcatcgccagcatttcgatcctcgt 146

RESULT 2
AAX01303
ID AAX01303 standard; DNA; 1041 BP.
XX
AC AAX01303;
XX
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (BioB) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS Escherichia coli.
XX
PN US5869719-A.
XX
PD 09-FEB-1999.
XX
PF 30-APR-1997; 97US-0846338.
XX
PR 30-APR-1997; 97US-0846338.
PR 08-MAR-1995; 95US-0401068.
XX

```
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Patton DA;
XX
DR   WPI; 1999-152902/13.
DR   P-PSDB; AAW73906.
XX
PT   Transgenic plants with high biotin levels - transformed with DNA
PT   encoding di:amino-pelargonic acid amino-transferase or biotin
PT   synthase
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.012;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcgt 25
          |||||||||||||||||||||||||
Db     90 gcatcgccagcatttcgatcctcgt 114


RESULT   3
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA;
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 0.012;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcgt 25
          |||||||||||||||||||||||||
Db    113 gcatcgccagcatttcgatcctcgt 137


RESULT   4
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzyme -
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 0.012;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcgt 25
          |||||||||||||||||||||||||
Db    131 gcatcgccagcatttcgatcctcgt 155


RESULT   5
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
```

```
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry)
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag= a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag= b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag= c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag= d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag= e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag= f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag= g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag= h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag= i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag= j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag= k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag= l
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag= m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag= n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                   7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag= o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag= p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag= q
FT                   /standard_name= "rho-independent transcriptional
FT                                   terminator"
FT   stem_loop       5583..5605
FT                   /*tag= r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;   93WO-EP02688.
XX
PR   02-OCT-1992;   92CH-0003124.
PR   15-JUL-1993;   93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.015;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||||||||||||||||||||||||
Db     206 gcatcgccagcatttcgatcctcgt 230


RESULT   6
AAV02308/c
ID   AAV02308 standard; cDNA; 3451 BP.
XX
AC   AAV02308;
```

```
XX
DT   06-MAY-1998  (first entry)
XX
DE   Cell membrane proton-ATPase encoding cDNA.
XX
KW   Cell membrane proton-ATPase; Cyanidium caldarium; acid resistance;
KW   transgenic plant; ds.
XX
OS   Cyanidium caldarium.
XX
FH   Key             Location/Qualifiers
FT   CDS             340..3207
FT                   /*tag= a
FT                   /product= "cell membrane proton-ATPase"
XX
PN   JP09252786-A.
XX
PD   30-SEP-1997.
XX
PF   25-MAR-1996;    96JP-0124124.
XX
PR   25-MAR-1996;    96JP-0124124.
XX
PA   (ORIY ) ORIENTAL YEAST CO LTD.
XX
DR   WPI; 1997-530157/49.
DR   P-PSDB; AAW31363.
XX
PT   Proton-ATPase and cDNA coding sequence - useful for producing
PT   acid-resistant transgenic plants
XX
PS   Claim 1; Page 11-13; 16pp; Japanese.
XX
CC   The present sequence encodes a cell membrane proton-ATPase.
CC   The gene is useful for producing acid-resistant transformed
CC   plants. It also allows large scale production of the enzyme.
XX
SQ   Sequence 3451 BP; 818 A; 860 C; 960 G; 813 T; 0 other;
```

```
  Query Match             72.8%;  Score 18.2;  DB 18;  Length 3451;
  Best Local Similarity   87.0%;  Pred. No. 24;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctc 23
           |||| |||||||||| || |||||
Db     812 GCATTGCCAGCATTGCGTTCCTC 790
```

```
RESULT   7
AAZ54339
ID   AAZ54339 standard; DNA; 645 BP.
XX
AC   AAZ54339;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 767 partial DNA sequence SEQ ID NO:2627.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;    99WO-US09346.
XX
PR   01-MAY-1998;    98US-0083758.
PR   31-JUL-1998;    98US-0094869.
PR   02-SEP-1998;    98US-0098994.
PR   02-SEP-1998;    98US-0099062.
PR   09-OCT-1998;    98US-0103749.
PR   09-OCT-1998;    98US-0103794.
PR   09-OCT-1998;    98US-0103796.
PR   25-FEB-1999;    99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   P-PSDB; AAY75577.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 7; Page 1246; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence 645 BP; 170 A; 138 C; 171 G; 166 T; 0 other;
```

```
  Query Match             71.2%;  Score 17.8;  DB 21;  Length 645;
  Best Local Similarity   90.5%;  Pred. No. 31;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcct 22
           ||| |||| ||||||||||||
Db     163 cattgccatcatttcgatcct 183
```

```
RESULT   8
AAZ54340
ID   AAZ54340 standard; DNA; 645 BP.
XX
AC   AAZ54340;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 767 partial DNA sequence SEQ ID NO:2629.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;    99WO-US09346.
XX
PR   01-MAY-1998;    98US-0083758.
PR   31-JUL-1998;    98US-0094869.
PR   02-SEP-1998;    98US-0098994.
```

```
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI   Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI   Tettelin H, Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   P-PSDB; AAY75578.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 7; Page 1247; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence 645 BP; 168 A; 147 C; 170 G; 160 T; 0 other;


  Query Match             71.2%;  Score 17.8;  DB 21;  Length 645;
  Best Local Similarity   90.5%;  Pred. No. 31;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcct 22
           ||| |||| ||||||||||||
Db     163 cattgccatcatttcgatcct 183


RESULT   9
AAZ54341
ID   AAZ54341 standard; DNA; 645 BP.
XX
AC   AAZ54341;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 767 partial DNA sequence SEQ ID NO:2631.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI   Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI   Tettelin H, Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   P-PSDB; AAY75579.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 7; Page 1247-1248; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence 645 BP; 170 A; 146 C; 170 G; 159 T; 0 other;


  Query Match             71.2%;  Score 17.8;  DB 21;  Length 645;
  Best Local Similarity   90.5%;  Pred. No. 31;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcct 22
           ||| |||| ||||||||||||
Db     163 cattgccatcatttcgatcct 183


RESULT   10
AAA81479/c
ID   AAA81479 standard; DNA; 69936 BP.
XX
AC   AAA81479;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_27 SEQ ID NO:27.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
```

PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;
XX
DR WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 547-567; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neissariae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 69936 BP; 16512 A; 19259 C; 17399 G; 16763 T; 3 other;

Query Match 71.2%; Score 17.8; DB 21; Length 69936;
Best Local Similarity 90.5%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 catcgccagcatttcgatcct 22
||| |||| ||||||||||||
Db 66524 CATTGCCATCATTTCGATCCT 66504

RESULT 11
AAF21607
ID AAF21607 standard; DNA; 349980 BP.
XX
AC AAF21607;
XX
DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:108.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX
OS Neisseria meningitidis.
XX
PN WO200066791-A1.
XX
PD 09-NOV-2000.
XX
PF 08-MAR-2000; 2000WO-US05928.
XX
PR 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX
DR WPI; 2000-647603/62.
XX
PT Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections -
XX
PS Claim 7; Appendix A; 692pp; English.
XX
CC The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX
SQ Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

Query Match 71.2%; Score 17.8; DB 21; Length 349980;
Best Local Similarity 90.5%; Pred. No. 65;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 catcgccagcatttcgatcct 22
||| |||| ||||||||||||
Db 114663 cattgccatcatttcgatcct 114683

RESULT 12
AAA81490
ID AAA81490 standard; DNA; 1437668 BP.
XX
AC AAA81490;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.

```
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 866-1272; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;


  Query Match             71.2%;  Score 17.8;  DB 21;  Length 1437668;
  Best Local Similarity   90.5%;  Pred. No. 73;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcct 22
           ||| |||| ||||||||||||
Db 414663 cattgccatcatttcgatcct 414683


RESULT  13
AAX13229/c
ID   AAX13229 standard; DNA; 20561 BP.
XX
AC   AAX13229;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:292.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Dillon PJ,  Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 1330-1340; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer-based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 20561 BP; 6898 A; 3503 C; 4035 G; 6112 T; 13 other;


  Query Match             70.4%;  Score 17.6;  DB 20;  Length 20561;
  Best Local Similarity   83.3%;  Pred. No. 58;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcgt 25
           |||| |||||||||  || |||||
Db    9218 CATCACCAGCATTTCGATACTCGT 9195


RESULT  14
AAX99364
ID   AAX99364 standard; DNA; 6386 BP.
XX
AC   AAX99364;
XX
DT   09-NOV-1999  (first entry)
XX
DE   Maize shrunken-1 (Sh1) gene sequence.
XX
KW   Gene expression; shrunken-1 gene; Sh1 gene; maize; foreign gene; ds.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   exon            131..182
FT                   /*tag=  a
FT                   /number= 1
FT   intron          183..1210
FT                   /*tag=  b
FT                   /number= 1
FT   exon            1211..1324
FT                   /*tag=  c
FT                   /number= 2
FT   intron          1325..1827
```

```
FT                   /*tag= d
FT                   /number= 2
FT   exon            1828..1948
FT                   /*tag= e
FT                   /number= 3
FT   intron          1949..2040
FT                   /*tag= f
FT                   /number= 3
FT   exon            2041..2187
FT                   /*tag= g
FT                   /number= 4
FT   intron          2188..2268
FT                   /*tag= h
FT                   /number= 4
FT   exon            2269..2460
FT                   /*tag= i
FT                   /number= 5
FT   intron          2461..2604
FT                   /*tag= j
FT                   /number= 5
FT   exon            2605..2728
FT                   /*tag= k
FT                   /number= 6
FT   intron          2729..2821
FT                   /*tag= l
FT                   /number= 6
FT   exon            2822..3038
FT                   /*tag= m
FT                   /number= 7
FT   intron          3039..3255
FT                   /*tag= n
FT                   /number= 7
FT   exon            3256..3351
FT                   /*tag= o
FT                   /number= 8
FT   intron          3352..3446
FT                   /*tag= p
FT                   /number= 8
FT   exon            3447..3620
FT                   /*tag= q
FT                   /number= 9
FT   intron          3621..3701
FT                   /*tag= r
FT                   /number= 9
FT   exon            3702..3818
FT                   /*tag= s
FT                   /number= 10
FT   intron          3819..3911
FT                   /*tag= t
FT                   /number= 10
FT   exon            3912..4078
FT                   /*tag= u
FT                   /number= 11
FT   intron          4079..4157
FT                   /*tag= v
FT                   /number= 11
FT   exon            4158..4381
FT                   /*tag= w
FT                   /number= 12
FT   intron          4382..4516
FT                   /*tag= x
FT                   /number= 12
FT   exon            4517..4835
FT                   /*tag= y
FT                   /number= 13
FT   intron          4836..4767
FT                   /*tag= z
FT                   /number= 13
FT   exon            4768..5212
FT                   /*tag= aa
FT                   /number= 14
FT   intron          5213..5371
FT                   /*tag= ab
FT                   /number= 14
FT   exon            5372..5510
FT                   /*tag= ac
FT                   /number= 15
FT   intron          5511..5635
FT                   /*tag= ad
FT                   /number= 15
FT   exon            5636..5917
FT                   /*tag= ae
FT                   /number= 16
XX
PN   US5955330-A.
XX
PD   21-SEP-1999.
XX
PF   07-JUN-1995;    95US-0483376.
XX
PR   18-MAY-1989;    89US-0353854.
PR   05-FEB-1992;    92US-0830956.
PR   04-AUG-1993;    93US-0102115.
PR   07-APR-1995;    95US-0418540.
PR   07-JUN-1995;    95US-0483376.
XX
PA   (RESE ) RESEARCH CORP TECHNOLOGIES INC.
XX
PI   Clancy MA,  Ferl RJ,  Hannah LC,  Vasil IK,  Vasil V;
XX
DR   WPI; 1999-539573/45.
XX
PT   Increasing the expression of a gene in a plant using an intron
XX
PS   Claim 6; Fig 1A-G; 32pp; English.
XX
CC   The invention relates to a method for increasing the expression rate of a
CC   gene in a plant cell using the first intron of the shrunken-1 (Sh1) locus
CC   of maize. The method comprises (a) inserting an intron into the gene
CC   comprising a 5' and 3' splice site of the Sh1 first intron, where the
CC   intron is inserted between a transcription start site and a translation
CC   start site of the gene to form an intron modified gene; and (b)
CC   introducing the modified gene into a plant cell so that the gene is
CC   expressed. The method is useful for increasing the expression of foreign
CC   genes in plant cells. The present sequence represents the maize Sh1
CC   gene sequence.
XX
SQ   Sequence 6386 BP; 1503 A; 1489 C; 1452 G; 1942 T; 0 other;


  Query Match             69.6%;  Score 17.4;  DB 20;  Length 6386;
  Best Local Similarity   94.7%;  Pred. No. 63;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatc 20
           ||||||| ||||||||||
Db    5698 catcgcctgcatttcgatc 5716


RESULT  15
AAZ19606
ID   AAZ19606 standard; RNA; 61 BP.
XX
AC   AAZ19606;
XX
DT   08-NOV-1999  (first entry)
XX
DE   Complement system protein C5 RNA binding ligand 13.
XX
KW   Complement system protein; C1q; C5; C3; RNA ligand; nootropic; stroke;
KW   neuroprotective; anti-asthmatic; cerebroprotective; immunosuppressive;
KW   diagnostic; prevention; treatment; complement protein-related disease;
KW   Alzheimer's disease; renal disease; transplant rejection; asthma; ss.
XX
OS   Synthetic.
XX
```

PN WO9941271-A1.
XX
PD 19-AUG-1999.
XX
PF 05-FEB-1999; 99WO-US02597.
XX
PR 29-SEP-1998; 98US-0163025.
PR 12-FEB-1998; 98US-0023228.
XX
PA (NEXS-) NEXSTAR PHARM INC.
XX
PI Biesecker G, Gold L;
XX
DR WPI; 1999-527357/44.
XX
PT New Nucleic Acid Ligand to complement protein C5, useful in the
PT treatment of Alzheimer's disease, renal diseases, transplant
PT rejection, stroke and asthma
XX
PS Example 4; Page 42; 120pp; English.
XX
CC This invention describes novel purified and isolated non-naturally
CC occurring RNA ligands (I) to complement system proteins C5, C1q and C3.
CC The products of the invention have nootropic, neuroprotective,
CC anti-asthmatic, cerebroprotective and immunosuppressive activity. The
CC new ligands are useful as diagnostic agents, and pharmaceutical agents
CC for prevention and treatment of complement protein-related diseases,
CC including Alzheimer's disease, renal diseases, transplant rejection,
CC stroke and asthma.
XX
SQ Sequence 61 BP; 11 A; 18 C; 21 G; 11 U; 0 other;

Query Match 67.2%; Score 16.8; DB 20; Length 61;
Best Local Similarity 65.0%; Pred. No. 69;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 5 cgccagcatttcgatcctcg 24
|| ||||||:::|||:| :||
Db 20 cgucagcauuucgaucuucg 39

RESULT 16
AAF61015
ID AAF61015 standard; DNA; 1497 BP.
XX
AC AAF61015;
XX
DT 16-MAY-2001 (first entry)
XX
DE P. putida KT2440-associated DNA ORF04209.
XX
KW Transgenic plant; detection; probe; amplification; vaccine carrier;
KW microbial production strain; biological remediation; ds.
XX
OS Pseudomonas putida.
XX
PN DE19935088-A1.
XX
PD 01-FEB-2001.
XX
PF 27-JUL-1999; 99DE-1035088.
XX
PR 27-JUL-1999; 99DE-1035088.
XX
PA (TIGR-) TIGR INST GENOMIC RES.
PA (QUIA-) QUIAGEN GMBH.
PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
DR WPI; 2001-192469/20.
XX
PT New DNA sequences specific for Pseudomonas putida KT2440, useful as
PT safe genetic engineering host, allow detection in presence of other
PT related bacteria -
XX
PS Claim 1a; Page 64-65; 158pp; German.
XX
CC This invention describes novel DNA sequences (I) for specific detection
CC of Pseudomonas putida KT2440. The invention also describes (1)
CC recombinant expression vector containing (I); (2) prokaryotic or
CC eukaryotic cells transformed or transfected with (I) or the vector of
CC (1); (3) production of expression products by culturing cells of (2);
CC (4) expression products, or their fragments, of (I) and synthetic
CC proteins or peptides with the same sequences (A); (5) poly- or
CC mono-clonal antibodies (Ab) that react specifically with (A); (6)
CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
CC plants that contain transformed or transfected cells of (2); (8)
CC detecting KT2440 using a labeled (I) or Ab as probe; and (9) DNA chips
CC carrying one or more (I). (I), and their fragments, are used as probes
CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
CC polymerase chain reaction, and for production of transgenic plants. (I),
CC or antibodies that recognize their expression products, are used for
CC detecting the presence of KT2440, particularly in presence of other,
CC even closely related, bacteria. KT2440 is one of the bacteria classified
CC as safe, by the National Institutes of Health, for genetic engineering
CC work, e.g. as microbial production strains, for biological remediation
CC and as vaccine carriers. (I) are exclusive to KT2440 with no significant
CC homology with sequences in other bacteria (specifically the closely
CC related pathogen P. aeruginosa). Compared with other 'safe' bacteria, it
CC has greater catabolic activity and better survival in, and adaptation to,
CC the rhizosphere and soil.
XX
SQ Sequence 1497 BP; 275 A; 515 C; 441 G; 264 T; 2 other;

Query Match 67.2%; Score 16.8; DB 22; Length 1497;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 cgccagcatttcgatcctcg 24
|||||||||| ||||||| ||
Db 360 cgccagcatgtcgatccccg 379

RESULT 17
AAC34251/c
ID AAC34251 standard; DNA; 609 BP.
XX
AC AAC34251;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 5983.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.

```
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             66.4%;  Score 16.6;  DB 21;  Length 609;
  Best Local Similarity   82.6%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctc 23
           ||||| |||||||| |||| ||
Db     159 GCATCACCAGCATTGAGATCATC 137


RESULT  18
AAF13337/c
ID   AAF13337 standard; cDNA; 671 BP.
XX
AC   AAF13337;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Aspergillus oryzae EST SEQ ID NO:5860.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS   Aspergillus oryzae.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999;   99US-0273623.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
PA   (NOVO ) NOVO NORDISK AS.
XX
PI   Berka RM,  Rey MW,  Shuster JR,  Kauppinen S,  Clausen IG,  Olsen PB;
XX
DR   WPI; 2000-594572/56.
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
PS   Claim 88; Page 2422-2423; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 671 BP; 164 A; 225 C; 157 G; 125 T; 0 other;

  Query Match             66.4%;  Score 16.6;  DB 21;  Length 671;
  Best Local Similarity   82.6%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 atcgccagcatttcgatcctcgt 25
           | || | |||||||| |||||||
Db     563 ACCGGCCGCATTTCGTTCCTCGT 541


RESULT  19
AAH82438
ID   AAH82438 standard; cDNA; 455 BP.
XX
AC   AAH82438;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human ovarian tumour associated polynucleotide sequence SEQ ID NO:62.
XX
KW   Human; ovarian tumour; ovarian cancer; diagnosis; gene therapy;
KW   immunogenic; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151513-A2.
XX
PD   19-JUL-2001.
XX
PF   16-JAN-2001; 2001WO-US01575.
XX
PR   14-JAN-2000; 2000US-0176722.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Algate PA;
XX
DR   WPI; 2001-425866/45.
XX
PT   Novel ovarian tumor proteins, and nucleic acids encoding them, used to
PT   treat and diagnose cancers, particularly ovarian cancer -
XX
PS   Claim 5; Page 72; 338pp; English.
XX
```

CC AAH82377 to AAH83878 represent human ovarian tumour-associated
CC polynucleotide sequences which encode ovarian tumour proteins. The
CC ovarian tumour protein and polynucleotide sequences have cytostatic
CC activity, and can be used in gene therapy and vaccine production. The
CC ovarian tumour proteins and polynucleotides can be used to inhibit
CC the development of cancer, particularly ovarian cancer. They can also
CC be used to diagnose the onset and progression of cancer.
XX
SQ Sequence 455 BP; 100 A; 101 C; 131 G; 101 T; 22 other;

Query Match 64.8%; Score 16.2; DB 22; Length 455;
Best Local Similarity 78.3%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcatcgccagcatttcgatcctc 23
|||| |||||||||| | |||
Db 290 gcatagccagcatttnagtnctc 312

RESULT 20
AAT79955/c
ID AAT79955 standard; DNA; 501 BP.
XX
AC AAT79955;
XX
DT 16-OCT-1997 (first entry)
XX
DE Fragment of cbl1A gene of Pseudomonas cepacia.
XX
KW Pseudomonas cepacia; transmissible lineage; Toronto/Edinburgh lineage;
KW cable adhesin type II PC pili; pilin protein; cystic fibrosis;
KW cblA gene; ss.
XX
OS Pseudomonas cepacia.
XX
PN WO9701647-A2.
XX
PD 16-JAN-1997.
XX
PF 28-JUN-1996; 96WO-US11132.
XX
PR 28-JUN-1995; 95US-0000828.
XX
PA (HEAL-) HEALTH & HOSPITALS CITY BOSTON.
XX
PI Goldstein R;
XX
DR WPI; 1997-100217/09.
XX
PT Identification of Pseudomonas cepacia lineages - using restriction
PT fragment length polymorphism analysis to identify highly
PT transmissible strains
XX
PS Disclosure; Fig 4; 52pp; English.
XX
CC This sequence represents nucleotides 1-501 of the cbl1A gene (encoding a
CC 17 kDa major subunit pilin protein) of Pseudomonas cepacia (PC). This
CC sequence is recognised by the primers and probes shown in
CC AAT79942-T79954. PC is an aerobic gram-negative bacillus with a
CC ubiquitous distribution in soil and water. It is an important pathogen
CC among cystic fibrosis (CF) patients, where patients infected by PC have a
CC higher morbidity and mortality than non-infected CF patients. The primers
CC and probes that recognise this sequence can be used in the method of the
CC invention. The method of the invention is for detecting the presence of a
CC strain of a transmissible lineage of PC in a sample. The method comprises
CC analysing the sample for restriction fragment length polymorphisms
CC (RFLPs) linked to a strain known to be of a transmissible lineage of PC.
CC Alternatively, the method comprises using one or more pairs of
CC oligonucleotide primers (such as AAT79942-T79945) having sequences
CC identical to portions of the gene encoding a 17 kDa major subunit pilin
CC protein of the cable adhesin type II PC pili. The methods are used for
CC identifying highly transmissible lineages of PC, especially the
CC Toronto/Edinburgh lineage. They are used particularly for studying the
CC pathogen in CF patients.
XX
SQ Sequence 501 BP; 109 A; 154 C; 153 G; 85 T; 0 other;

Query Match 64.8%; Score 16.2; DB 18; Length 501;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 cgccagcatttcgatcctcgt 25
|| |||||||||||| | ||||
Db 114 CGACAGCATTTCGAGCGTCGT 94

RESULT 21
AAC53704/c
ID AAC53704 standard; DNA; 543 BP.
XX
AC AAC53704;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 75424.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 64.8%; Score 16.2; DB 21; Length 543;

```
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcgccagcatttcgatcctc 23
           |||| || |||||||||| ||
Db     122 ATCGGCAACATTTCGATCTTC 102


RESULT  22
AAF12787/c
ID   AAF12787 standard; cDNA; 675 BP.
XX
AC   AAF12787;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Aspergillus oryzae EST SEQ ID NO:5310.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS   Aspergillus oryzae.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999;   99US-0273623.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
PA   (NOVO ) NOVO NORDISK AS.
XX
PI   Berka RM,  Rey MW,  Shuster JR,  Kauppinen S,  Clausen IG,  Olsen PB;
XX
DR   WPI; 2000-594572/56.
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
PS   Claim 88; Page 2215; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 675 BP; 171 A; 183 C; 172 G; 143 T; 6 other;

  Query Match             64.8%;  Score 16.2;  DB 21;  Length 675;
  Best Local Similarity   85.7%;  Pred. No. 1.8e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcc 21
           ||||||||| |||| |||| |
Db     248 GCATCGCCACCATTCCGATGC 228


RESULT  23
AAC54459/c
ID   AAC54459 standard; DNA; 1451 BP.
XX
AC   AAC54459;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 77945.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
```

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 64.8%; Score 16.2; DB 21; Length 1451;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcgccagcatttcgatcctc 23

```
          |||| || |||||||||| ||
Db    417 ATCGGCAACATTTCGATCTTC 397


RESULT  24
AAC54360/c
ID   AAC54360 standard; DNA; 1452 BP.
XX
AC   AAC54360;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 77596.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
```

```
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             64.8%;  Score 16.2;  DB 21;  Length 1452;
  Best Local Similarity   85.7%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcgccagcatttcgatcctc 23
           |||| || ||||||||||| ||
Db     417 ATCGGCAACATTTCGATCTTC 397


RESULT  25
AAC39263/c
ID   AAC39263 standard; DNA; 1546 BP.

XX
AC   AAC39263;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 23979.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
```

PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 64.8%; Score 16.2; DB 21; Length 1546;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       3 atcgccagcatttcgatcctc 23
           |||| || |||||||||| ||
Db     511 ATCGGCAACATTTCGATCTTC 491
```

RESULT 26
AAC54458/c
ID AAC54458 standard; DNA; 1549 BP.
XX
AC AAC54458;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 77941.
XX

KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

```
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             64.8%;  Score 16.2;  DB 21;  Length 1549;
  Best Local Similarity   85.7%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcgccagcatttcgatcctc 23
           |||| || ||||||||||  ||
Db     511 ATCGGCAACATTTCGATCTTC 491


RESULT  27
AAZ55887
ID   AAZ55887 standard; DNA; 68750 BP.
XX
AC   AAZ55887;
XX
DT   10-APR-2000  (first entry)
XX
DE   Sorangium cellulosum 68.75 kb contig.
XX
KW   Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW   anticancer; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1826
FT                   /*tag=  a
FT                   /partial
FT                   /product= "Partial Orf 1 protein (AAY58580)"
FT                   /note= "No initiation codon given in the specification"
FT   CDS             complement (1900..3171)
FT                   /*tag=  b
FT                   /product= "Orf 2 protein (AAY58581)"
FT   CDS             3415..5556
FT                   /*tag=  c
FT                   /product= "Orf 3 protein (AAY58582)"
FT   CDS             complement (5612..5992)
FT                   /*tag=  d
FT                   /product= "Orf 4 protein (AAY58583)"
FT   CDS             6226..6675
FT                   /*tag=  e
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag=  f
FT                   /product= "Type I polyketide synthase, EPOS A
FT   (AAY58573)" CDS              11872..116104
FT                   /*tag=  g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                   (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag=  h
FT                   /product= "Type I polyketide synthase, EPOS B
FT   (AAY58575)" CDS              21746..43519
FT                   /*tag=  i
FT                   /product= "Type I polyketide synthase, EPOS C
FT   (AAY58576)" CDS              43524..54920
FT                   /*tag=  j
FT                   /product= "Type I polyketide synthase, EPOS D
FT   (AAY58577)" CDS              54935..62254
FT                   /*tag=  k
FT                   /product= "Type I polyketide synthase, EPOS E
FT   (AAY58578)" CDS              62369..63628
FT                   /*tag=  l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                   (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag=  m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag=  n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag=  o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag=  p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag=  q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag=  r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag=  s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag=  t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag=  u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag=  v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
```

XX
PD 23-DEC-1999.
XX
PF 16-JUN-1999; 99WO-EP04171.
XX
PR 18-JUN-1998; 98US-0099504.
PR 24-SEP-1998; 98US-0101631.
PR 05-FEB-1999; 99US-0118906.
XX
PA (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX
DR WPI; 2000-097741/08.
DR P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR AAY58592, AAY58593, AAY58594.
XX
PT New isolated epothilone synthase genes, used for the recombinant
PT production of epothilone for use in cancer therapy -
XX
PS Claim 14; Page 87-104; 174pp; English.
XX
CC This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC comprising 22 open reading frames (ORFs) and includes genes encoding
CC proteins involved in the biosynthesis of epothilones. Epothilones A and
CC B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC starter unit; polyketides being synthesised from two-carbon building
CC blocks, the beta-carbon of which always carries a keto group. Each round
CC of two-carbon addition is carried out by a complex of enzymes known as
CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC involved in transport. Epothilones mimic the biological activity of
CC taxol, and may be substituted for taxol in cancer chemotherapeutic
CC compositions. Epothilones exhibit a much lower drop in potency against a
CC multiply drug-resistant cell line compared with taxol, and are
CC considerably less efficiently exported from such cells by the multidrug
CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC epothilones as anticancer agents, they are problematical to produce on a
CC large scale. Epothilones are too complex for industrial scale chemical
CC synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC poor yields of epothilones. The nucleic acids of the invention may be
CC used for the recombinant production of epothilones in a heterologous host
CC that is more amenable to fermentation.
XX
SQ Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

Query Match 64.8%; Score 16.2; DB 21; Length 68750;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 cgccagcatttcgatcctcgt 25
||| ||| |||||| ||||||
Db 41953 cgcgagcctttcgaccctcgt 41973

RESULT 28
AAH42540
ID AAH42540 standard; DNA; 58 BP.
XX
AC AAH42540;
XX
DT 01-OCT-2001 (first entry)
XX
DE Oligonucleotide used to construct a probe for CAR sequences.
XX
KW Constitutive androstane receptor; CAR; CAR-mediated disorder;
KW cholesterol indicator; cholesterol level; hypercholesterolaemia;
KW lipid disorder; atherosclerosis; cardiovascular disorder; ss.
XX
OS Mus sp.
XX
PN WO200151045-A2.
XX
PD 19-JUL-2001.
XX
PF 12-JAN-2001; 2001WO-US01111.
XX
PR 13-JAN-2000; 2000US-0176398.
XX
PA (TULA-) TULARIK INC.
XX
PI Lehmann JM, Shiau AK;
XX
DR WPI; 2001-476019/51.
XX
PT Identifying agent for treating CAR-mediated disorder, involves
PT screening agent that modulates CAR-mediated intermolecular interaction
PT and determining if the agent modulates cholesterol level in test mammal
PT -
XX
PS Example 1; Page 43; 76pp; English.
XX
CC The specification describes a method for identifying an agent for
CC treating constitutive androstane receptor (CAR)-mediated disorder/
CC condition. The method comprises screening candidate agents that
CC modulate CAR-mediated intermolecular interaction and determining if
CC administration of the agent to test mammal modulates cholesterol
CC indicator level, or determining change in cholesterol indicator level
CC in the CAR compromised mammal to which the agent is administered
CC compared to control. Agents which modulate CAR-mediated regulation of
CC cholesterol levels, are useful for treating a CAR-mediated disorder or
CC condition such as hypercholesterolaemia, lipid disorders, atherosclerosis
CC and cardiovascular disorders. The agent is useful for testing potential
CC CAR-mediated disorder treatments, and for studying mechanisms of
CC hypercholesterolaemia. Overlapping oligonucleotides AAH42539-40 were
CC used to construct a probe, which was used to identify ES cells expressing
CC CAR sequences.
XX
SQ Sequence 58 BP; 8 A; 25 C; 11 G; 14 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 58;
Best Local Similarity 79.2%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcatcgccagcatttcgatcctcg 24
||||| || || |||| | |||||
Db 29 gcatcaccggcctttccagcctcg 52

RESULT 29
AAA67164/c
ID AAA67164 standard; DNA; 205 BP.
XX
AC AAA67164;
XX
DT 31-OCT-2000 (first entry)
XX
DE Pinus radiata 1,3-beta-D-glucanase DNA sequence SEQ ID NO:165.
XX
KW Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant; ds.
XX
OS Pinus radiata.
XX
PN WO200022092-A2.
XX

```
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant -
XX
PS   Claim 1; Page 118; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 205 BP; 57 A; 55 C; 42 G; 51 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 21;  Length 205;
  Best Local Similarity   79.2%;  Pred. No. 1.9e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcgt 25
           |||| ||||| ||||  | |||||
Db     137 CATCACCAGCTTTTCTGTTCTCGT 114


RESULT  30
AAA67162/c
ID   AAA67162 standard; DNA; 297 BP.
XX
AC   AAA67162;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Pinus radiata 1,3-beta-D-glucanase DNA sequence SEQ ID NO:163.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Pinus radiata.
XX
PN   WO200022092-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant -
XX
PS   Claim 1; Page 117; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 297 BP; 73 A; 80 C; 69 G; 75 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 21;  Length 297;
  Best Local Similarity   79.2%;  Pred. No. 2e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcgt 25
           |||| ||||| ||||  | |||||
Db     243 CATCACCAGCTTTTCTGTTCTCGT 220


RESULT  31
AAA67165/c
ID   AAA67165 standard; DNA; 393 BP.
XX
AC   AAA67165;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Pinus radiata 1,3-beta-D-glucanase DNA sequence SEQ ID NO:166.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Pinus radiata.
XX
PN   WO200022092-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant -
XX
PS   Claim 1; Page 118; 301pp; English.
XX
```

```
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 393 BP; 95 A; 106 C; 91 G; 101 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 21;  Length 393;
  Best Local Similarity   79.2%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcgt 25
           |||| ||||| |||| | |||||
Db     317 CATCACCAGCTTTTCTGTTCTCGT 294


RESULT  32
AAA67166/c
ID   AAA67166 standard; DNA; 394 BP.
XX
AC   AAA67166;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Pinus radiata 1,3-beta-D-glucanase DNA sequence SEQ ID NO:167.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Pinus radiata.
XX
PN   WO200022092-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant  -
XX
PS   Claim 1; Page 118; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 394 BP; 95 A; 106 C; 91 G; 102 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 21;  Length 394;
  Best Local Similarity   79.2%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcgt 25
           |||| ||||| |||| | |||||
Db     317 CATCACCAGCTTTTCTGTTCTCGT 294


RESULT  33
AAA67357/c
ID   AAA67357 standard; DNA; 399 BP.
XX
AC   AAA67357;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Pinus radiata endo-1,3-1,4-beta glucanase DNA sequence SEQ ID NO:358.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Pinus radiata.
XX
PN   WO200022092-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant  -
XX
PS   Claim 1; Page 166; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 399 BP; 96 A; 108 C; 93 G; 102 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 21;  Length 399;
  Best Local Similarity   79.2%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;
```

```
Qy       2 catcgccagcatttcgatcctcgt 25
           |||| ||||| ||||  | |||||
Db     317 CATCACCAGCTTTTCTGTTCTCGT 294


RESULT  34
AAA67169/c
ID   AAA67169 standard; DNA; 419 BP.
XX
AC   AAA67169;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Pinus radiata 1,3-beta-D-glucanase DNA sequence SEQ ID NO:170.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Pinus radiata.
XX
PN   WO200022092-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant  -
XX
PS   Claim 1; Page 119; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 419 BP; 104 A; 114 C; 98 G; 103 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 21;  Length 419;
  Best Local Similarity   79.2%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcgt 25
           |||| ||||| ||||  | |||||
Db     317 CATCACCAGCTTTTCTGTTCTCGT 294


RESULT  35
AAA67358/c
ID   AAA67358 standard; DNA; 469 BP.
XX
AC   AAA67358;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Pinus radiata endo-1,3-1,4-beta glucanase DNA sequence SEQ ID NO:359.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Pinus radiata.
XX
PN   WO200022092-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-NZ00169.
XX
PR   13-OCT-1998;   98US-0170862.
PR   11-AUG-1999;   99US-0148426.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Bloksberg LN;
XX
DR   WPI; 2000-339328/29.
XX
PT   New genes encoding proteins involved in a plant polysaccharide
PT   biosynthetic pathway, useful for modulating or altering the
PT   polysaccharide content, composition or structure of the plant  -
XX
PS   Claim 1; Page 166; 301pp; English.
XX
CC   The present invention describes isolated polynucleotides (PN) comprising
CC   a sequence selected from one of 835 nucleotide sequences given in
CC   AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC   an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC   sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC   835 sequences or sequences that are degenerately equivalent or allelic
CC   to the 835 sequences. The polynucleotides are used to modify the
CC   activity of a polypeptide involved in a polysaccharide biosynthetic
CC   pathway in the plant. They are especially used to modulate or alter the
CC   polysaccharide content, composition or structure of the plant. AAB16268
CC   to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC   given in the present invention.
XX
SQ   Sequence 469 BP; 119 A; 122 C; 105 G; 123 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 21;  Length 469;
  Best Local Similarity   79.2%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcgt 25
           |||| ||||| ||||  | |||||
Db     317 CATCACCAGCTTTTCTGTTCTCGT 294


RESULT  36
AAA67167/c
ID   AAA67167 standard; DNA; 498 BP.
XX
AC   AAA67167;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Pinus radiata 1,3-beta-D-glucanase DNA sequence SEQ ID NO:168.
XX
KW   Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW   plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW   transgenic plant; ds.
XX
OS   Pinus radiata.
```

XX
PN WO200022092-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-NZ00169.
XX
PR 13-OCT-1998; 98US-0170862.
PR 11-AUG-1999; 99US-0148426.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Bloksberg LN;
XX
DR WPI; 2000-339328/29.
XX
PT New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant -
XX
PS Claim 1; Page 119; 301pp; English.
XX
CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX
SQ Sequence 498 BP; 126 A; 129 C; 110 G; 130 T; 3 other;

Query Match 64.0%; Score 16; DB 21; Length 498;
Best Local Similarity 79.2%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 catcgccagcatttcgatcctcgt 25
|||| ||||| |||| | |||||
Db 316 CATCACCAGCTTTTCTGTTCTCGT 293

RESULT 37
AAH66006
ID AAH66006 standard; DNA; 699 BP.
XX
AC AAH66006;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 1041.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR P-PSDB; AAG90787.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 8; SEQ ID NO: 1041; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 699 BP; 149 A; 182 C; 210 G; 158 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 699;
Best Local Similarity 79.2%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcatcgccagcatttcgatcctcg 24
||| || || | | ||||||||||
Db 77 gcaacggcatcctgtcgatcctcg 100

RESULT 38
AAF67973
ID AAF67973 standard; DNA; 822 BP.
XX
AC AAF67973;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:461.
XX
KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100805-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00926.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 09-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.

```
PR   09-JUL-1999;   99DE-1032180.
PR   09-JUL-1999;   99DE-1032182.
PR   09-JUL-1999;   99DE-1032190.
PR   09-JUL-1999;   99DE-1032191.
PR   09-JUL-1999;   99DE-1032209.
PR   09-JUL-1999;   99DE-1032212.
PR   09-JUL-1999;   99DE-1032227.
PR   09-JUL-1999;   99DE-1032228.
PR   09-JUL-1999;   99DE-1032229.
PR   09-JUL-1999;   99DE-1032230.
PR   14-JUL-1999;   99DE-1032927.
PR   14-JUL-1999;   99DE-1033005.
PR   14-JUL-1999;   99DE-1033006.
PR   27-AUG-1999;   99DE-1040764.
PR   27-AUG-1999;   99DE-1040765.
PR   27-AUG-1999;   99DE-1040766.
PR   27-AUG-1999;   99DE-1040830.
PR   27-AUG-1999;   99DE-1040831.
PR   27-AUG-1999;   99DE-1040832.
PR   27-AUG-1999;   99DE-1040833.
PR   31-AUG-1999;   99DE-1041378.
PR   31-AUG-1999;   99DE-1041379.
PR   31-AUG-1999;   99DE-1041395.
PR   03-SEP-1999;   99DE-1042077.
PR   03-SEP-1999;   99DE-1042078.
PR   03-SEP-1999;   99DE-1042079.
PR   03-SEP-1999;   99DE-1042088.
XX
PA   (BADI ) BASF AG.
XX
PI   Pompejus M,  Kroeger B,  Schroeder H,  Zelder O,  Haberhauer G;
XX
DR   WPI; 2001-071486/08.
DR   P-PSDB; AAB76740.
XX
PT   Corynebacterium glutamicum nucleic acids encoding membrane construction
PT   and membrane transport proteins or their portions, useful for typing or
PT   identifying C. glutamicum or related bacteria, and as markers for
PT   transformation -
XX
PS   Claim 3; Page 807-808; 1119pp; English.
XX
CC   AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC   construction and membrane transport (MCT) proteins given in AAB76510 to
CC   AAB76847. The MCT nucleic acids and proteins are useful in the
CC   identification of microorganisms which can be used to produce fine
CC   chemicals, for modulating fine chemical production in C. glutamicum or
CC   related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC   identification of C. glutamicum or related bacteria, as reference points
CC   for mapping C. glutamicum genome, and as markers for transformation.
CC   AAF68082 and AAF68082 represent sequencing primers which are used in an
CC   example from the present invention.
XX
SQ   Sequence 822 BP; 174 A; 202 C; 254 G; 192 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 22;  Length 822;
  Best Local Similarity   79.2%;  Pred. No. 2.3e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcg 24
           ||| || || | | ||||||||||
Db     177 gcaacggcatcctgtcgatcctcg 200


RESULT  39
AAF12760
ID   AAF12760 standard; cDNA; 868 BP.
XX
AC   AAF12760;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Aspergillus oryzae EST SEQ ID NO:5283.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS   Aspergillus oryzae.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999;   99US-0273623.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
PA   (NOVO ) NOVO NORDISK AS.
XX
PI   Berka RM,  Rey MW,  Shuster JR,  Kauppinen S,  Clausen IG,  Olsen PB;
XX
DR   WPI; 2000-594572/56.
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
PS   Claim 88; Page 2205; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 868 BP; 186 A; 245 C; 192 G; 245 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 21;  Length 868;
  Best Local Similarity   79.2%;  Pred. No. 2.3e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcg 24
           | |||||| || ||| |||||||
Db     143 gtatcgccggcttttacatcctcg 166


RESULT  40
AAH48556
ID   AAH48556 standard; DNA; 943 BP.
XX
AC   AAH48556;
```

```
XX
DT   20-SEP-2001  (first entry)
XX
DE   Human fascin DNA fragment SEQ ID 6.
XX
KW   Fascin; regulatory sequence; human; dendritic cell; antiviral; tumor;
KW   antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
KW   immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
KW   Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
KW   autoimmune disease; transplant rejection; ds.
XX
OS   Homo sapiens.
XX
PN   WO200151631-A2.
XX
PD   19-JUL-2001.
XX
PF   12-JAN-2001; 2001WO-EP00362.
XX
PR   13-JAN-2000; 2000DE-1001169.
PR   02-MAR-2000; 2000DE-1010188.
XX
PA   (RESK/) RESKE-KUNZ A.
PA   (ROSS/) ROSS X.
PA   (ROSS/) ROSS R.
PA   (BROS/) BROS M.
XX
PI   Reske-Kunz A,  Ross X,  Ross R,  Bros M;
XX
DR   WPI; 2001-451858/48.
XX
PT   New regulatory sequences from the fascin gene, useful for providing
PT   dendritic cell-specific expression of e.g. antigens, e.g. for
PT   vaccination against tumors and infections  -
XX
PS   Claim 1c; Page 87; 117pp; German.
XX
CC   This invention describes novel regulatory sequences (A) derived from
CC   human fascin that provide specific expression in dendritic cells (DC) and
CC   which have antiviral, antibacterial, antifungal, antiparasitic,
CC   anti-allergic, neurological, immunomodulatory and apoptotic activity.
CC   (A) are used to regulate expression of antigens, immunoregulators,
CC   antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
CC   and host cells that contain (A) are useful: (i) in vaccines against
CC   viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
CC   Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
CC   tumors, allergies, infections, autoimmune diseases and transplant
CC   rejection. They can also be provide specific expression of antigens and
CC   immunoregulators in DC; for isolation and identification of cell factors
CC   and cis-elements from regulatory sequences that mediate DC-specific
CC   expression; to determine the degree of maturity of DC and to block
CC   transcription factors, by providing binding sites in DC. (A) provide
CC   DC-specific expression of nucleic acid under their control, allowing a
CC   more specific regulation of the immune response and eliminating the long
CC   and laborious purification of DC (since a complete leucocyte population
CC   may be transformed), including transformation in vitro. This sequence
CC   represents a DNA fragment of human fascin which is described in the
CC   invention.
XX
SQ   Sequence 943 BP; 135 A; 307 C; 340 G; 161 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 22;  Length 943;
  Best Local Similarity   79.2%;  Pred. No. 2.3e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcg 24
           |||||||||||||||  ||   |||
Db     111 gcatcgccagcattgggaatgtcg 134


RESULT  41
AAC36656/c

ID   AAC36656 standard; DNA; 1135 BP.
XX
AC   AAC36656;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 14601.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
```

PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 64.0%; Score 16; DB 21; Length 1135;
Best Local Similarity 79.2%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy      1 gcatcgccagcatttcgatcctcg 24
          ||||| |||||||||  || || |
Db    426 GCATCACCAGCATTTTTATTCTTG 403
```

RESULT 42
AAH48555
ID AAH48555 standard; DNA; 1240 BP.
XX
AC AAH48555;
XX
DT 20-SEP-2001 (first entry)
XX
DE Human fascin DNA fragment SEQ ID 5.

```
XX
KW   Fascin; regulatory sequence; human; dendritic cell; antiviral; tumor;
KW   antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
KW   immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
KW   Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
KW   autoimmune disease; transplant rejection; ds.
XX
OS   Homo sapiens.
XX
PN   WO200151631-A2.
XX
PD   19-JUL-2001.
XX
PF   12-JAN-2001; 2001WO-EP00362.
XX
PR   13-JAN-2000; 2000DE-1001169.
PR   02-MAR-2000; 2000DE-1010188.
XX
PA   (RESK/) RESKE-KUNZ A.
PA   (ROSS/) ROSS X.
PA   (ROSS/) ROSS R.
PA   (BROS/) BROS M.
XX
PI   Reske-Kunz A,  Ross X,  Ross R,  Bros M;
XX
DR   WPI; 2001-451858/48.
XX
PT   New regulatory sequences from the fascin gene, useful for providing
PT   dendritic cell-specific expression of e.g. antigens, e.g. for
PT   vaccination against tumors and infections -
XX
PS   Claim 1c; Page 87; 117pp; German.
XX
CC   This invention describes novel regulatory sequences (A) derived from
CC   human fascin that provide specific expression in dendritic cells (DC) and
CC   which have antiviral, antibacterial, antifungal, antiparasitic,
CC   anti-allergic, neurological, immunomodulatory and apoptotic activity.
CC   (A) are used to regulate expression of antigens, immunoregulators,
CC   antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
CC   and host cells that contain (A) are useful: (i) in vaccines against
CC   viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
CC   Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
CC   tumors, allergies, infections, autoimmune diseases and transplant
CC   rejection. They can also be provide specific expression of antigens and
CC   immunoregulators in DC; for isolation and identification of cell factors
CC   and cis-elements from regulatory sequences that mediate DC-specific
CC   expression; to determine the degree of maturity of DC and to block
CC   transcription factors, by providing binding sites in DC. (A) provide
CC   DC-specific expression of nucleic acid under their control, allowing a
CC   more specific regulation of the immune response and eliminating the long
CC   and laborious purification of DC (since a complete leucocyte population
CC   may be transformed), including transformation in vitro. This sequence
CC   represents a DNA fragment of human fascin which is described in the
CC   invention.
XX
SQ   Sequence 1240 BP; 191 A; 364 C; 466 G; 219 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 22;  Length 1240;
  Best Local Similarity   79.2%;  Pred. No. 2.4e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcg 24
           |||||||||||||  ||   |||
Db     408 gcatcgccagcattgggaatgtcg 431


RESULT  43
AAX80215
ID   AAX80215 standard; cDNA; 1280 BP.
XX
AC   AAX80215;
XX
DT   18-AUG-1999  (first entry)
XX
DE   Mouse nuclear receptor protein nNR4 encoding cDNA.
XX
KW   Mouse; nuclear receptor protein; nNR4; identification; differentiation;
KW   cell proliferation; regulation; murine; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             155..1228
FT                   /*tag=  a
XX
PN   WO9929722-A1.
XX
PD   17-JUN-1999.
XX
PF   11-DEC-1998;   98WO-US26446.
XX
PR   12-DEC-1997;   97US-0068144.
XX
PA   (MERI ) MERCK & CO INC.
XX
PI   Chen F;
XX
DR   WPI; 1999-385573/32.
DR   P-PSDB; AAY17872.
XX
PT   Novel DNA encoding murine nuclear receptor
XX
PS   Claim 9; Page 30-31; 47pp; English.
XX
CC   The present sequence encodes a mouse nuclear receptor protein designated
CC   nNR4. The nNR4 protein is useful in the identification of downstream
CC   target genes and ligands regulating its activity. The nuclear receptor
CC   is involved in the regulation of in vivo cell proliferation and/or cell
CC   development. The nNR4 polynucleotides, expression vectors and host cells
CC   are useful for the recombinant production of the protein.
XX
SQ   Sequence 1280 BP; 309 A; 348 C; 343 G; 280 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 20;  Length 1280;
  Best Local Similarity   79.2%;  Pred. No. 2.4e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcg 24
           |||||  ||  ||  ||||  |  |||||
Db     613 gcatcaccggcctttccagcctcg 636


RESULT  44
AAX24003
ID   AAX24003 standard; DNA; 1361 BP.
XX
AC   AAX24003;
XX
DT   29-JUN-1999  (first entry)
XX
DE   Mouse CAR receptor DNA.
XX
KW   CAR receptor; constitutive activator of retinoic acid response element;
KW   cytostatic; antithyroid; nuclear hormone receptor superfamily; cancer;
KW   zinc finger transcription factor; retinoic acid; treatment; lung;
KW   Grave's disease; thyroid hormone receptor; mouse; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             148..1224
FT                   /*tag=  a
FT                   /product= "CAR receptor"
XX
```

PN WO9915555-A1.
XX
PD 01-APR-1999.
XX
PF 17-SEP-1998; 98WO-US19365.
XX
PR 19-SEP-1997; 97US-0934388.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Baes MI, Choi H, Moore DD;
XX
DR WPI; 1999-254691/21.
DR P-PSDB; AAW93903.
XX
PT CAR (Constitutive Activator of Retinoic Acid Response Elements)
PT receptor polypeptides, used to identify ligands for therapy of
PT Grave's disease and lung cancer
XX
PS Claim 2; Fig 2; 66pp; English.
XX
CC This invention describes the isolation of novel human and mouse CAR
CC (Constitutive Activator of Retinoic Acid Response Elements) receptor
CC polypeptides which have cytostatic and antithyroid activity. The CAR
CC receptor polypeptides are members of the nuclear hormone receptor
CC superfamily (zinc finger transcription factors). The CAR receptor
CC polypeptides bind to their target DNA sequence and activate expression of
CC downstream genes, even in the absence of retinoic acid. CAR receptor
CC ligands are useful for treating Grave's disease by decreasing thyroid
CC hormone receptor function, and for treating cancer (especially lung
CC cancer) by increasing retinoic acid receptor expression.
XX
SQ Sequence 1361 BP; 344 A; 362 C; 364 G; 291 T; 0 other;

Query Match 64.0%; Score 16; DB 20; Length 1361;
Best Local Similarity 79.2%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcatcgccagcatttcgatcctcg 24
||||| || || |||| | |||||
Db 606 gcatcaccggcctttccagcctcg 629

RESULT 45
AAH48554
ID AAH48554 standard; DNA; 1449 BP.
XX
AC AAH48554;
XX
DT 20-SEP-2001 (first entry)
XX
DE Human fascin DNA fragment SEQ ID 4.
XX
KW Fascin; regulatory sequence; human; dendritic cell; antiviral; tumor;
KW antibacterial; antifungal; antiparasitic; anti-allergic; neurological;
KW immunomodulatory; apoptotic; expression regulator; vaccine; allergen;
KW Creutzfeld-Jakob disease; Alzheimer's disease; gene therapy;
KW autoimmune disease; transplant rejection; ds.
XX
OS Homo sapiens.
XX
PN WO200151631-A2.
XX
PD 19-JUL-2001.
XX
PF 12-JAN-2001; 2001WO-EP00362.
XX
PR 13-JAN-2000; 2000DE-1001169.
PR 02-MAR-2000; 2000DE-1010188.
XX
PA (RESK/) RESKE-KUNZ A.
PA (ROSS/) ROSS X.
PA (ROSS/) ROSS R.
PA (BROS/) BROS M.
XX
PI Reske-Kunz A, Ross X, Ross R, Bros M;
XX
DR WPI; 2001-451858/48.
XX
PT New regulatory sequences from the fascin gene, useful for providing
PT dendritic cell-specific expression of e.g. antigens, e.g. for
PT vaccination against tumors and infections -
XX
PS Claim 1c; Page 86-87; 117pp; German.
XX
CC This invention describes novel regulatory sequences (A) derived from
CC human fascin that provide specific expression in dendritic cells (DC) and
CC which have antiviral, antibacterial, antifungal, antiparasitic,
CC anti-allergic, neurological, immunomodulatory and apoptotic activity.
CC (A) are used to regulate expression of antigens, immunoregulators,
CC antisense sequences etc. in DC-specific fashion. Recombinant DNA, vectors
CC and host cells that contain (A) are useful: (i) in vaccines against
CC viruses, bacteria, fungi, parasites, tumors, allergens and plaques in
CC Creutzfeld-Jakob and Alzheimer's disease; and (ii) for gene therapy of
CC tumors, allergies, infections, autoimmune diseases and transplant
CC rejection. They can also be provide specific expression of antigens and
CC immunoregulators in DC; for isolation and identification of cell factors
CC and cis-elements from regulatory sequences that mediate DC-specific
CC expression; to determine the degree of maturity of DC and to block
CC transcription factors, by providing binding sites in DC. (A) provide
CC DC-specific expression of nucleic acid under their control, allowing a
CC more specific regulation of the immune response and eliminating the long
CC and laborious purification of DC (since a complete leucocyte population
CC may be transformed), including transformation in vitro. This sequence
CC represents a DNA fragment of human fascin which is described in the
CC invention.
XX
SQ Sequence 1449 BP; 232 A; 429 C; 532 G; 256 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 1449;
Best Local Similarity 79.2%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcatcgccagcatttcgatcctcg 24
||||||||||||| || |||
Db 617 gcatcgccagcattgggaatgtcg 640

Search completed: December 26, 2001, 12:19:25
Job time: 6926 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:25:34 ; Search time 141.92 Seconds
(without alignments)
39.895 Million cell updates/sec

Title: US-09-396-196F-7
Perfect score: 25
Sequence: 1 gcatcgccagcatttcgatcctcgt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
	2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
	3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
	4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
	5	17.4	69.6	6386	2	US-08-483-376-1	Sequence 1, Appli
	6	16.8	67.2	61	3	US-09-023-228B-59	Sequence 59, Appl
	7	16.2	64.8	851	4	US-08-998-416-294	Sequence 294, App
	8	16.2	64.8	12732	4	US-09-060-756-1	Sequence 1, Appli
	9	16.2	64.8	68750	3	US-09-335-409-1	Sequence 1, Appli
	10	16	64.0	9960	3	US-08-822-586-46	Sequence 46, Appl
	11	16	64.0	10095	3	US-08-822-586-45	Sequence 45, Appl
	12	16	64.0	176373	3	US-09-128-155-17	Sequence 17, Appl
	13	15.6	62.4	978	3	US-08-706-281A-17	Sequence 17, Appl
	14	15.6	62.4	978	4	US-09-097-231-17	Sequence 17, Appl
	15	15.6	62.4	1147	1	US-08-417-103-15	Sequence 15, Appl
	16	15.6	62.4	1351	1	US-07-816-283-5	Sequence 5, Appli
	17	15.6	62.4	1351	1	US-08-417-103-5	Sequence 5, Appli
	18	15.6	62.4	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	19	15.6	62.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	20	15.4	61.6	723	4	US-08-998-416-617	Sequence 617, App
c	21	15.4	61.6	1126	2	US-08-479-733A-27	Sequence 27, Appl
c	22	15.4	61.6	1126	3	US-08-487-427-27	Sequence 27, Appl
c	23	15.4	61.6	1126	3	US-08-479-727A-27	Sequence 27, Appl
c	24	15.4	61.6	1126	3	US-08-482-369A-27	Sequence 27, Appl
c	25	15.4	61.6	1126	5	PCT-US95-07439-27	Sequence 27, Appl
c	26	15.4	61.6	1404	4	US-09-202-101-15	Sequence 15, Appl
c	27	15.4	61.6	1500	1	US-08-487-037-4	Sequence 4, Appli
c	28	15.2	60.8	1712	2	US-08-632-598-1	Sequence 1, Appli
c	29	15.2	60.8	1712	4	US-09-231-240-1	Sequence 1, Appli
c	30	15.2	60.8	1725	4	US-09-380-061B-19	Sequence 19, Appl
c	31	15.2	60.8	1956	2	US-08-762-308-10	Sequence 10, Appl
c	32	15.2	60.8	2099	1	US-08-299-849B-25	Sequence 25, Appl
c	33	15.2	60.8	2099	2	US-08-142-368A-25	Sequence 25, Appl
c	34	15.2	60.8	2099	3	US-08-967-727-25	Sequence 25, Appl
c	35	15.2	60.8	2099	4	US-08-037-230D-25	Sequence 25, Appl
c	36	15.2	60.8	2150	1	US-08-299-849B-24	Sequence 24, Appl
c	37	15.2	60.8	2150	2	US-08-142-368A-24	Sequence 24, Appl
c	38	15.2	60.8	2150	3	US-08-967-727-24	Sequence 24, Appl
c	39	15.2	60.8	2150	4	US-08-037-230D-24	Sequence 24, Appl
	40	15.2	60.8	2190	4	US-09-625-188-19	Sequence 19, Appl
c	41	15.2	60.8	3029	3	US-08-693-214-1	Sequence 1, Appli
c	42	15.2	60.8	4724	1	US-08-404-665-3	Sequence 3, Appli
c	43	15.2	60.8	4724	1	US-08-404-671-3	Sequence 3, Appli
c	44	15.2	60.8	4724	1	US-08-404-781-3	Sequence 3, Appli
c	45	15.2	60.8	87350	3	US-08-781-891-79	Sequence 79, Appl
	46	15	60.0	102	1	US-08-247-475-24	Sequence 24, Appl
	47	15	60.0	102	1	US-08-479-650-24	Sequence 24, Appl
	48	15	60.0	102	1	US-08-191-866D-30	Sequence 30, Appl
	49	15	60.0	102	1	US-08-674-169-24	Sequence 24, Appl
	50	15	60.0	102	2	US-08-185-949B-30	Sequence 30, Appl
c	51	15	60.0	918	1	US-08-328-322-3	Sequence 3, Appli
c	52	15	60.0	1222	1	US-08-328-322-1	Sequence 1, Appli
	53	15	60.0	1406	2	US-08-933-750C-72	Sequence 72, Appl
	54	15	60.0	1406	3	US-09-234-613-72	Sequence 72, Appl
c	55	15	60.0	1907	3	US-08-826-611-1	Sequence 1, Appli
c	56	15	60.0	2581	1	US-08-200-900A-1	Sequence 1, Appli
c	57	15	60.0	2581	5	PCT-US94-00616-1	Sequence 1, Appli
c	58	15	60.0	2696	1	US-07-961-522-1	Sequence 1, Appli
c	59	15	60.0	2696	1	US-08-217-438-1	Sequence 1, Appli
c	60	15	60.0	2696	1	US-08-321-978-1	Sequence 1, Appli
c	61	15	60.0	2696	2	US-08-710-584-1	Sequence 1, Appli
c	62	15	60.0	2745	1	US-08-328-322-6	Sequence 6, Appli
	63	15	60.0	2774	3	US-08-732-433-2	Sequence 2, Appli
c	64	15	60.0	3500	1	US-08-328-322-4	Sequence 4, Appli
	65	15	60.0	4773	3	US-08-884-324-9	Sequence 9, Appli
	66	15	60.0	5517	4	US-09-120-663-1	Sequence 1, Appli
	67	15	60.0	8931	3	US-09-051-019-1	Sequence 1, Appli
	68	15	60.0	9515	1	US-08-920-812-13	Sequence 13, Appl
	69	15	60.0	9515	1	US-08-920-827-13	Sequence 13, Appl
	70	15	60.0	9515	1	US-08-921-177-13	Sequence 13, Appl
	71	15	60.0	9515	1	US-08-362-577C-13	Sequence 13, Appl
	72	15	60.0	9515	2	US-08-920-828-13	Sequence 13, Appl
	73	15	60.0	11464	3	US-08-884-324-13	Sequence 13, Appl
	74	15	60.0	12752	2	US-08-459-146-1	Sequence 1, Appli
	75	15	60.0	12752	2	US-08-459-065-1	Sequence 1, Appli
	76	15	60.0	28994	3	US-08-884-324-14	Sequence 14, Appl
	77	14.8	59.2	1779	5	PCT-US96-05320A-831	Sequence 831, App
	78	14.8	59.2	3606	1	US-07-661-610C-7	Sequence 7, Appli
c	79	14.8	59.2	6831	2	US-08-609-049A-27	Sequence 27, Appl
c	80	14.8	59.2	6831	4	US-09-170-996-27	Sequence 27, Appl
	81	14.6	58.4	27	3	US-08-513-974B-161	Sequence 161, App
c	82	14.6	58.4	600	4	US-09-385-982-402	Sequence 402, App
	83	14.6	58.4	706	4	US-08-797-812-25	Sequence 25, Appl
	84	14.6	58.4	1395	4	US-08-806-263-1	Sequence 1, Appli
c	85	14.6	58.4	1444	1	US-07-814-964-9	Sequence 9, Appli
c	86	14.6	58.4	1444	1	US-08-258-442-9	Sequence 9, Appli
c	87	14.6	58.4	1444	1	US-08-328-809-4	Sequence 4, Appli
c	88	14.6	58.4	1444	5	PCT-US92-11107-9	Sequence 9, Appli
c	89	14.6	58.4	1668	4	US-09-385-028-20	Sequence 20, Appl
	90	14.6	58.4	1881	2	US-08-519-081-1	Sequence 1, Appli
	91	14.6	58.4	1881	2	US-08-754-369-1	Sequence 1, Appli
c	92	14.6	58.4	1898	1	US-07-814-964-8	Sequence 8, Appli
c	93	14.6	58.4	1898	1	US-08-258-442-8	Sequence 8, Appli
c	94	14.6	58.4	1898	1	US-08-328-809-3	Sequence 3, Appli
c	95	14.6	58.4	1898	5	PCT-US92-11107-8	Sequence 8, Appli
	96	14.6	58.4	2082	2	US-08-785-310A-2	Sequence 2, Appli
c	97	14.6	58.4	2111	1	US-08-528-255A-2	Sequence 2, Appli
c	98	14.6	58.4	2111	1	US-08-717-365-2	Sequence 2, Appli
	99	14.6	58.4	2266	2	US-08-724-394A-18	Sequence 18, Appl
c	100	14.6	58.4	2839	1	US-07-814-964-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT   1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Enhanced Biotin Biosynthesis in Plant Tissue
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/351,970
;      FILING DATE:  08-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-401-068-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.0067;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||||||||||||||||||||||||
Db      90 GCATCGCCAGCATTTCGATCCTCGT 114


RESULT   2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-846-338-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.0067;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||||||||||||||||||||||||
Db      90 GCATCGCCAGCATTTCGATCCTCGT 114


RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A-15/9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1157
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 117
;      OTHER INFORMATION:  /product= "Biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioB"
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2295..3050
;      OTHER INFORMATION:  /codon_start= 2295
;      OTHER INFORMATION:  /function= "involved in pimeloyl-CoA synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /gene= "bioC"
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3750..5039
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3750
;      OTHER INFORMATION:  /EC_number= 2.6.1.62
;      OTHER INFORMATION:  /product= "DAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioA"
;      OTHER INFORMATION:  /number= 5
;      OTHER INFORMATION:  /standard_name=
;      OTHER INFORMATION:  "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
;      OTHER INFORMATION:  aminotransf."
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  5098..5574
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 5098
;      OTHER INFORMATION:  /function= "unknown, involved in biotin synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "ORFI"
;      OTHER INFORMATION:  /number= 6
;    FEATURE:
;      NAME/KEY:  -10_signal
;      LOCATION:  45..49
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  -35_signal
;      LOCATION:  23..28
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  105..119
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "bioB RBS no.9"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  2284..2297
;      OTHER INFORMATION:  /standard_name= "bioC RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3742..3752
;      OTHER INFORMATION:  /standard_name= "bioA RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  5088..5100
;      OTHER INFORMATION:  /standard_name= "ORFI RBS"
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  5583..5644
;      OTHER INFORMATION:  /standard_name= "rho-independent
;      OTHER INFORMATION:  transcriptional terminator"
;    FEATURE:
;      NAME/KEY:  stem_loop
;      LOCATION:  5583..5605
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..96
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "promoter ptac"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-1


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0081;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||||||||||||||||||||||||
Db     206 GCATCGCCAGCATTTCGATCCTCGT 230


RESULT   4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A15-9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1154..2308
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1154
;      OTHER INFORMATION:  /EC_number= 2.3.1.47
;      OTHER INFORMATION:  /product= "KAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "8-Amino-7-oxononanoate synthase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3043..3753
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3043
;      OTHER INFORMATION:  /EC_number= 6.3.3.3
;      OTHER INFORMATION:  /product= "DTB synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioD"
;      OTHER INFORMATION:  /number= 4
;      OTHER INFORMATION:  /standard_name= "Dethiobiotin synthase"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1141..1156
;      OTHER INFORMATION:  /standard_name= "bioF RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3030..3045
;      OTHER INFORMATION:  /standard_name= "bioD RBS"
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-6


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0081;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||||||||||||||||||||||||
Db     206 GCATCGCCAGCATTTCGATCCTCGT 230


RESULT   5
US-08-483-376-1
; Sequence 1, Application US/08483376
; Patent No. 5955330
;  GENERAL INFORMATION:
;    APPLICANT:  Vasil, Vimla
;    APPLICANT:  Clancy, Maureen A.
;    APPLICANT:  Ferl, Robert J.
;    APPLICANT:  Vasil, Indra K.
;    APPLICANT:  Hannah, L. C.
;    TITLE OF INVENTION:  No. 5955330el Means for Enhancing Gene
;    TITLE OF INVENTION:  Expression
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,376
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/418,540
;      FILING DATE:  07-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/102,115
;      FILING DATE:  04-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/830,956
;      FILING DATE:  05-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/353,854
;      FILING DATE:  18-MAY-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  10-94B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6386 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Zea mays
;      STRAIN:  Black Sweet
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  131..182
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1211..1324
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1828..1948
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  2041..2187
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  2269..2460
;    FEATURE:
;      NAME/KEY:  exon
```

```
;      LOCATION:  2605..2728
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  2822..3038
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  3256..3351
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  3447..3620
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  3702..3818
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  3912..4078
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  4158..4381
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  4517..4835
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  4768..5212
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  5372..5510
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  5636..5917
US-08-483-376-1


  Query Match             69.6%;  Score 17.4;  DB 2;  Length 6386;
  Best Local Similarity   94.7%;  Pred. No. 25;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatc 20
           ||||||| ||||||||||||
Db    5698 CATCGCCTGCATTTCGATC 5716


RESULT   6
US-09-023-228B-59
; Sequence 59, Application US/09023228B
; Patent No. 6140490
;  GENERAL INFORMATION:
;    APPLICANT:  BIESECKER, GREGORY
;    APPLICANT:  GOLD, LARRY
;    TITLE OF INVENTION:  HIGH AFFINITY NUCLEIC ACID LIGANDS OF
;    TITLE OF INVENTION:  COMPLEMENT SYSTEM PROTEINS
;    NUMBER OF SEQUENCES:  157
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Swanson & Bratschun, L.L.C.
;      STREET:  8400 E. Prentice Place #200
;      CITY:  Denver
;      STATE:  Colorado
;      COUNTRY:  US
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 8.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/023,228B
;      FILING DATE:  12-FEBRUARY-1998
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US97/01739
;      FILING DATE:  30 JAN 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/595,335
;      FILING DATE:  1 FEB 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barry J. Swanson, Esq.
;      REGISTRATION NUMBER:  33,215
;      REFERENCE/DOCKET NUMBER:  NEX50/CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 793-3333
;      TELEFAX:  (303) 793-3433
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  61 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  RNA
;    FEATURE:
;      OTHER INFORMATION:  All pyrimidines are 2'-F modified
US-09-023-228B-59


  Query Match             67.2%;  Score 16.8;  DB 3;  Length 61;
  Best Local Similarity   65.0%;  Pred. No. 28;
  Matches   13;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 cgccagcatttcgatcctcg 24
           || |||||:::|||:| :||
Db      20 CGUCAGCAUUUCGAUCUUCG 39


RESULT   7
US-08-998-416-294
; Sequence 294, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  294:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  851 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1244RP
US-08-998-416-294


  Query Match             64.8%;  Score 16.2;  DB 4;  Length 851;
  Best Local Similarity   85.7%;  Pred. No. 71;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      5 cgccagcatttcgatcctcgt 25
          ||||||||||| | || ||||
Db    503 CGCCAGCATTTTGTTCATCGT 523


RESULT   8
US-09-060-756-1
; Sequence 1, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
;  APPLICANT: Cole, Stewart
;  APPLICANT: Buchrieser-Brosch, Roland
;  APPLICANT: Gordon, Stephen
;  APPLICANT: Billault, Alain
;  TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
;  TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
;  TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
;  FILE REFERENCE: 3495-0169
;  CURRENT APPLICATION NUMBER: US/09/060,756
;  CURRENT FILING DATE: 1998-04-16
;  NUMBER OF SEQ ID NOS: 743
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 12732
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-09-060-756-1


  Query Match             64.8%;  Score 16.2;  DB 4;  Length 12732;
  Best Local Similarity   85.7%;  Pred. No. 95;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcc 21
          ||| |||||||||  ||||||
Db  11641 gcagcgccagcatggcgatcc 11661


RESULT   9
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT: Ligon, James
;  APPLICANT: Molnar, Istvan
;  APPLICANT: Zirkle, Ross
;  APPLICANT: Cyr, Devon
;  APPLICANT: Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/335,409
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-335-409-1


  Query Match             64.8%;  Score 16.2;  DB 3;  Length 68750;
  Best Local Similarity   85.7%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      5 cgccagcatttcgatcctcgt 25
          ||| ||| |||||| ||||||
Db  41953 cgcgagcctttcgaccctcgt 41973


RESULT  10
US-08-822-586-46
; Sequence 46, Application US/08822586
; Patent No. 6015890
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAM R. JACOBS, JR., JAMES M. MUSSER AND
;    APPLICANT:  AMALIO TELENTI
;    TITLE OF INVENTION:  AN EMBCAB OPERON OF MYCOBACTERIA AND
;    TITLE OF INVENTION:  MUTANTS THEREOF
;    NUMBER OF SEQUENCES:  57
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  AMSTER, ROTHSTEIN & EBENSTEIN
;      STREET:  90 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH 1.44 Mb STORAGE
;      MEDIUM TYPE:  DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/822,586
;      FILING DATE:  MARCH 20, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ELIZABETH A. BOGOSIAN
;      REGISTRATION NUMBER:  39,911
;      REFERENCE/DOCKET NUMBER:  96700/437
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  46:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9960
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    HYPOTHETICAL:  NO
US-08-822-586-46


  Query Match             64.0%;  Score 16;  DB 3;  Length 9960;
  Best Local Similarity   79.2%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcg 24
          ||||| | | ||| ||| ||||||
Db   1784 GCATCACGATCATCTCGTTCCTCG 1807


RESULT  11
US-08-822-586-45
; Sequence 45, Application US/08822586
; Patent No. 6015890
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAM R. JACOBS, JR., JAMES M. MUSSER AND
;    APPLICANT:  AMALIO TELENTI
;    TITLE OF INVENTION:  AN EMBCAB OPERON OF MYCOBACTERIA AND
```

```
;    TITLE OF INVENTION:  MUTANTS THEREOF
;    NUMBER OF SEQUENCES:  57
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  AMSTER, ROTHSTEIN & EBENSTEIN
;      STREET:  90 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH 1.44 Mb STORAGE
;      MEDIUM TYPE:  DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/822,586
;      FILING DATE:  MARCH 20, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ELIZABETH A. BOGOSIAN
;      REGISTRATION NUMBER:  39,911
;      REFERENCE/DOCKET NUMBER:  96700/437
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10095
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    HYPOTHETICAL:  NO
US-08-822-586-45


  Query Match             64.0%;  Score 16;  DB 3;  Length 10095;
  Best Local Similarity   79.2%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcg 24
           ||||| | | |||||| ||||||
Db    1859 GCATCACGATCATTTCCTTCCTCG 1882


RESULT  12
US-09-128-155-17
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
;   LENGTH: 176373
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(176373)
;   OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17


  Query Match             64.0%;  Score 16;  DB 3;  Length 176373;
  Best Local Similarity   79.2%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctcgt 25
           |||| ||||| |||| || || ||
Db   97026 catctccagcctttctattcttgt 97049


RESULT  13
US-08-706-281A-17
; Sequence 17, Application US/08706281A
; Patent No. 6100048
;  GENERAL INFORMATION:
;    APPLICANT:  Cone, Roger D
;    APPLICANT:  Fan, Wei
;    APPLICANT:  Boston, Bruce A
;    APPLICANT:  Kesterton, Robert A
;    APPLICANT:  Lu, Dongsi
;    APPLICANT:  Chen, Wenbiao
;    TITLE OF INVENTION:  Methods and Reagents for Discovering and
;    TITLE OF INVENTION:  Using Mammalian Melanocortin Receptor Agonists and Antagoni
;    TITLE OF INVENTION:  To Modulate Feeding Behavior in Animals
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
;      STREET:  300 South Wacker Drive
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/706,281A
;      FILING DATE:  04-SEP-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6100048nan, Kevin E
;      REGISTRATION NUMBER:  35,303
;      REFERENCE/DOCKET NUMBER:  96,886
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-913-0001
;      TELEFAX:  312-913-0002
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  978 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..975
US-08-706-281A-17


  Query Match             62.4%;  Score 15.6;  DB 3;  Length 978;
  Best Local Similarity   81.8%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctc 23
           ||||||| |||||| || | ||
Db     486 CATCGCCTGCATTTGGACCTTC 507


RESULT  14
US-09-097-231-17
```

```
; Sequence 17, Application US/09097231
; Patent No. 6278038
;   GENERAL INFORMATION:
;        APPLICANT: Cone, Roger D
;                   Chen, Wenbiao
;                   Low, Malcolm J
;        TITLE OF INVENTION: Mammalian Melanocortin Receptor and Uses
;        NUMBER OF SEQUENCES: 22
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
;             STREET: 300 South Wacker Drive
;             CITY: Chicago
;             STATE: Illinois
;             COUNTRY: USA
;             ZIP: 60606
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/097,231
;             FILING DATE: 12-Jun-1998
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: No. 6278038nan, Kevin E
;             REGISTRATION NUMBER: 35,303
;             REFERENCE/DOCKET NUMBER: 96,886-C
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 312-913-0001
;             TELEFAX: 312-913-0002
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 17:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 978 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..975
;        SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-097-231-17


  Query Match             62.4%;  Score 15.6;  DB 4;  Length 978;
  Best Local Similarity   81.8%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctc 23
           ||||||| |||||| || | ||
Db     486 CATCGCCTGCATTTGGACCTTC 507


RESULT  15
US-08-417-103-15
; Sequence 15, Application US/08417103
; Patent No. 5723299
;  GENERAL INFORMATION:
;     APPLICANT:  Bell, Graeme I.
;     APPLICANT:  Yamada, Yuichiro
;     APPLICANT:  Seino, Susumu
;     TITLE OF INVENTION:  SOMATOSTATIN RECEPTORS
;     NUMBER OF SEQUENCES:  16
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold, White & Durkee
;       STREET:  P.O. Box 4433
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  United States of America
;       ZIP:  77210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/417,103
;       FILING DATE:  05-APR-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/816,283
;       FILING DATE:  01-DEC-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Wilson, Mark B.
;       REGISTRATION NUMBER:  37,259
;       REFERENCE/DOCKET NUMBER:  ARCD:144
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (512) 418-3000
;       TELEFAX:  (512) 474-7577
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1147 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  15..1121
US-08-417-103-15


  Query Match             62.4%;  Score 15.6;  DB 1;  Length 1147;
  Best Local Similarity   81.8%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcctc 23
           |||| ||| |||||  ||||||
Db     242 CATCACCAACATTTACATCCTC 263


RESULT  16
US-07-816-283-5
; Sequence 5, Application US/07816283
; Patent No. 5436155
;  GENERAL INFORMATION:
;     APPLICANT:  Bell, Graeme I.
;     APPLICANT:  Yamada, Yuichiro
;     APPLICANT:  Seino, Susumu
;     TITLE OF INVENTION:  SOMATOSTATIN RECEPTORS
;     NUMBER OF SEQUENCES:  12
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold, White & Durkee
;       STREET:  PO Box 4433
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  USA
;       ZIP:  77210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/816,283
;       FILING DATE:  19911231
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  McDaniel, C. Steven
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  713-787-1400
;       TELEFAX:  713-789-2679
;       TELEX:  79-0924
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  1351 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-07-816-283-5


  Query Match             62.4%;  Score 15.6;  DB 1;  Length 1351;
  Best Local Similarity   81.8%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catcgccagcatttcgatcctc 23
          |||| ||| |||||  ||||||
Db    310 CATCACCAACATTTACATCCTC 331


RESULT  17
US-08-417-103-5
; Sequence 5, Application US/08417103
; Patent No. 5723299
;  GENERAL INFORMATION:
;    APPLICANT:  Bell, Graeme I.
;    APPLICANT:  Yamada, Yuichiro
;    APPLICANT:  Seino, Susumu
;    TITLE OF INVENTION:  SOMATOSTATIN RECEPTORS
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  United States of America
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/417,103
;      FILING DATE:  05-APR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/816,283
;      FILING DATE:  01-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wilson, Mark B.
;      REGISTRATION NUMBER:  37,259
;      REFERENCE/DOCKET NUMBER:  ARCD:144
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1351 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  83..1189
US-08-417-103-5


  Query Match             62.4%;  Score 15.6;  DB 1;  Length 1351;
  Best Local Similarity   81.8%;  Pred. No. 1.4e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catcgccagcatttcgatcctc 23
          |||| ||| |||||  ||||||
Db    310 CATCACCAACATTTACATCCTC 331


RESULT  18
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


  Query Match             62.4%;  Score 15.6;  DB 4;  Length 4403765;
  Best Local Similarity   85.7%;  Pred. No. 51;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcc 21
          ||| ||||||||||  ||||||
Db 1707784 gcagcgccagcatggcgatcc 1707804


RESULT  19
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1


  Query Match             62.4%;  Score 15.6;  DB 4;  Length 4411529;
  Best Local Similarity   85.7%;  Pred. No. 51;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcc 21
          ||| ||||||||||  ||||||
Db 1707655 gcagcgccagcatggcgatcc 1707675


RESULT  20
US-08-998-416-617
; Sequence 617, Application US/08998416
```

```
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  617:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  723 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1419RP
US-08-998-416-617


  Query Match             61.6%;  Score 15.4;  DB 4;  Length 723;
  Best Local Similarity   76.0%;  Pred. No. 1.6e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           ||| ||| ||||||  ||| | |||
Db     263 GCACCGCGAGCATTCTGATACACGT 287


RESULT  21
US-08-479-733A-27/c
; Sequence 27, Application US/08479733A
; Patent No. 5877289
;  GENERAL INFORMATION:
;    APPLICANT:  Thorpe, Philip E.
;    APPLICANT:  Edgington, Thomas S.
;    TITLE OF INVENTION:  Methods and Compositions for the
;    TITLE OF INVENTION:  Specific Coagulation of Vasculature
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  US
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,733A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/273,567
;      FILING DATE:  11-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:459/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  512/418-3000
;      TELEFAX:  512/474-7577
;      TELEX:  N/A
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1126 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-479-733A-27


  Query Match             61.6%;  Score 15.4;  DB 2;  Length 1126;
  Best Local Similarity   76.0%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           ||| | |||||||| | || ||| ||
Db     796 GCACCTCCAGCATCTTGAGCCTGGT 772


RESULT  22
US-08-487-427-27/c
; Sequence 27, Application US/08487427
; Patent No. 6004555
;  GENERAL INFORMATION:
;    APPLICANT:  Thorpe, Philip E.
;    APPLICANT:  Edgington, Thomas S.
;    TITLE OF INVENTION:  Methods and Compositions for the
;    TITLE OF INVENTION:  Specific Coagulation of Vasculature
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  US
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,427
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/273,567
;      FILING DATE:  11-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
```

```
;      REFERENCE/DOCKET NUMBER:  UTSD:457/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  512/418-3000
;      TELEFAX:  512/474-7577
;      TELEX:  N/A
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1126 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-487-427-27


  Query Match             61.6%;  Score 15.4;  DB 3;  Length 1126;
  Best Local Similarity   76.0%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcgt 25
          ||| | ||||||| | || ||| ||
Db    796 GCACCTCCAGCATCTTGAGCCTGGT 772


RESULT  23
US-08-479-727A-27/c
; Sequence 27, Application US/08479727A
; Patent No. 6036955
;  GENERAL INFORMATION:
;    APPLICANT:  Thorpe, Philip E.
;    TITLE OF INVENTION:  Methods and Compositions for the
;    TITLE OF INVENTION:  Specific Coagulation of Vasculature
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  US
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,727A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/273,567
;      FILING DATE:  11-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:461/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  512/418-3000
;      TELEFAX:  512/474-7577
;      TELEX:  N/A
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1126 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-479-727A-27


  Query Match             61.6%;  Score 15.4;  DB 3;  Length 1126;
  Best Local Similarity   76.0%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcgt 25
          ||| | ||||||| | || ||| ||
Db    796 GCACCTCCAGCATCTTGAGCCTGGT 772


RESULT  24
US-08-482-369A-27/c
; Sequence 27, Application US/08482369A
; Patent No. 6093399
;  GENERAL INFORMATION:
;    APPLICANT:  Thorpe, Philip E.
;    APPLICANT:  Edgington, Thomas S.
;    TITLE OF INVENTION:  Methods and Compositions for the Specific
;    TITLE OF INVENTION:  Coagulation of Vasculature
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  US
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/482,369A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/273,567
;      FILING DATE:  11-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:433/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  512/418-3000
;      TELEFAX:  512/474-7577
;      TELEX:  N/A
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1126 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-482-369A-27


  Query Match             61.6%;  Score 15.4;  DB 3;  Length 1126;
  Best Local Similarity   76.0%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcgt 25
          ||| | ||||||| | || ||| ||
Db    796 GCACCTCCAGCATCTTGAGCCTGGT 772


RESULT  25
PCT-US95-07439-27/c
; Sequence 27, Application PC/TUS9507439
;  GENERAL INFORMATION:
;    APPLICANT:
;    APPLICANT:  NAME:     BOARD OF REGENTS, THE UNIVERSITY OF
;    APPLICANT:  TEXAS SYSTEM
;    APPLICANT:  STREET:   201 West 7th Street
;    APPLICANT:  CITY:     Austin
;    APPLICANT:  STATE:    Texas
;    APPLICANT:  COUNTRY:  United States of America
;    APPLICANT:  POSTAL CODE:  78701
;    APPLICANT:  TELEPHONE NO: (512)499-4462
;    APPLICANT:  TELEFAX:  (512)499-4523
```

```
;     APPLICANT:  NAME:      THE SCRIPPS RESEARCH INSTITUTE
;     APPLICANT:  STREET:    10666 North Torrey Pines Road
;     APPLICANT:  CITY:      LaJolla
;     APPLICANT:  STATE:     California
;     APPLICANT:  COUNTRY:   United States of America
;     APPLICANT:  POSTAL CODE:  92037
;     TITLE OF INVENTION:  METHODS AND COMPOSITIONS
;     TITLE OF INVENTION:  FOR THE SPECIFIC
;     TITLE OF INVENTION:  COAGULATION OF VASCULATURE
;     NUMBER OF SEQUENCES:  27
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold, White & Durkee
;       STREET:  P. O. Box 4433
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  USA
;       ZIP:  77210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS, ASCII
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/07439
;       FILING DATE:  Concurrently herewith
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/273,567
;       FILING DATE:  11-JUN-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PARKER, DAVID L.
;       REGISTRATION NUMBER:  32,165
;       REFERENCE/DOCKET NUMBER:  UTFD433P--
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (512) 418-3000
;       TELEFAX:  (713) 789-2679
;       TELEX:  79-0924
;   INFORMATION FOR SEQ ID NO:  27:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1126 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
PCT-US95-07439-27


  Query Match             61.6%;  Score 15.4;  DB 5;  Length 1126;
  Best Local Similarity   76.0%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           ||| | ||||||| | || ||| ||
Db     796 GCACCTCCAGCATCTTGAGCCTGGT 772


RESULT  26
US-09-202-101-15/c
; Sequence 15, Application US/09202101
; Patent No. 6277618
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Recombinant blood-coagulation proteases
;    NUMBER OF SEQUENCES:  17
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/202,101
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1404 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-202-101-15


  Query Match             61.6%;  Score 15.4;  DB 4;  Length 1404;
  Best Local Similarity   76.0%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           ||| | ||||||| | || ||| ||
Db    1084 GCACCTCCAGCATCTTGAGCCTGGT 1060


RESULT  27
US-08-487-037-4/c
; Sequence 4, Application US/08487037
; Patent No. 5795863
;  GENERAL INFORMATION:
;    APPLICANT:  Wolf, David L.
;    TITLE OF INVENTION:  RECOMBINANT AGENTS AFFECTING THROMBOSIS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Avenue, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,037
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adler, Reid G.
;      REGISTRATION NUMBER:  30,988
;      REFERENCE/DOCKET NUMBER:  2803-0002.02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;      TELEX:  90-4030
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1500 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-487-037-4


  Query Match             61.6%;  Score 15.4;  DB 1;  Length 1500;
  Best Local Similarity   76.0%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           ||| | ||||||| | || ||| ||
Db    1179 GCACCTCCAGCATCTTGAGCCTGGT 1155


RESULT  28
US-08-632-598-1/c
; Sequence 1, Application US/08632598
; Patent No. 5886164
;  GENERAL INFORMATION:
;    APPLICANT:  BIRD, COLIN R
;    APPLICANT:  FLETCHER, JONATHON D
```

```
;    TITLE OF INVENTION:  RIPENING-RELATED GENES FROM BANANA
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY AND CUSHMAN
;      STREET:  1100 NEW YORK AVENUE N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/632,598
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  223355/SEE50112/US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  861-3000
;      TELEFAX:  822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1712 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  MUSA
;    IMMEDIATE SOURCE:
;      CLONE:  ACS GENE
US-08-632-598-1


  Query Match             60.8%;  Score 15.2;  DB 2;  Length 1712;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       6 gccagcatttcgatcctcgt 25
           || |||| ||| ||||||||
Db    1669 GCAAGCAATTCTATCCTCGT 1650


RESULT  29
US-09-231-240-1/c
; Sequence 1, Application US/09231240
; Patent No. 6262346
;  GENERAL INFORMATION:
;    APPLICANT:  BIRD, COLIN R
;    APPLICANT:  FLETCHER, JONATHON D
;    TITLE OF INVENTION:  RIPENING-RELATED GENES FROM BANANA
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY AND CUSHMAN
;      STREET:  1100 NEW YORK AVENUE N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/231,240
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/632,598
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  223355/SEE50112/US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  861-3000
;      TELEFAX:  822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1712 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  MUSA
;    IMMEDIATE SOURCE:
;      CLONE:  ACS GENE
US-09-231-240-1


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 1712;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       6 gccagcatttcgatcctcgt 25
           || |||| ||| ||||||||
Db    1669 GCAAGCAATTCTATCCTCGT 1650


RESULT  30
US-09-380-061B-19/c
; Sequence 19, Application US/09380061B
; Patent No. 6265177
;   GENERAL INFORMATION:
;        APPLICANT: SQUIRRELL, DAVID JAMES
;                   WHITE, PETER JOHN
;                   LOWE, CHRISTOPHER ROBIN
;                   MURRAY, JAMES AUGUSTUS HENRY
;        TITLE OF INVENTION: ENZYME ASSAY FOR MUTANT FIREFLY LUCIFERASE
;        NUMBER OF SEQUENCES: 21
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: NIXON & VANDERHYE P.C.
;             STREET: 1100 NORTH GLEBE ROAD
;             CITY: ARLINGTON
;             STATE: VIRGINIA
;             COUNTRY: U.S.A.
;             ZIP: 22201-4714
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/380,061B
;             FILING DATE: 25-Aug-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB98/01026
;             FILING DATE: 7-APR-1998
;             APPLICATION NUMBER: GB 9707468.8
;             FILING DATE: 11-APR-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: SADOFF, B. J.
;             REGISTRATION NUMBER: 36,663
;             REFERENCE/DOCKET NUMBER: 124-725
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703)816-4000
```

```
;             TELEFAX: (703)816-4100
;    INFORMATION FOR SEQ ID NO: 19:
;         SEQUENCE CHARACTERISTICS:
;             LENGTH: 1725 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: both
;             TOPOLOGY: linear
;         MOLECULE TYPE: cDNA
;         FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  32..1675
;         SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-380-061B-19


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 1725;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 tcgccagcatttcgatcctc 23
           || |||||| ||| ||||||
Db     102 TCTCCAGCAGTTCCATCCTC 83


RESULT  31
US-08-762-308-10/c
; Sequence 10, Application US/08762308
; Patent No. 5925548
;  GENERAL INFORMATION:
;    APPLICANT:  Beutler, Bruce A.
;    APPLICANT:  Bazzoni, Flavia M.
;    TITLE OF INVENTION:  MODIFIED RECEPTORS THAT CONTINUOUSLY
;    TITLE OF INVENTION:  SIGNAL
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  TX
;      COUNTRY:  USA
;      ZIP:  77210-4433
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/762,308
;      FILING DATE:  09-DEC-1996
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/224,593
;      FILING DATE:  05-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kitchell, Barbara S.
;      REGISTRATION NUMBER:  33,928
;      REFERENCE/DOCKET NUMBER:  UTSD:335--1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  418-3000
;      TELEFAX:  474-7577
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1956 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-762-308-10


  Query Match             60.8%;  Score 15.2;  DB 2;  Length 1956;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatc 20
           ||||| ||||| | ||||||
Db    1266 GCATCTCCAGCCTCTCGATC 1247


RESULT  32
US-08-299-849B-25/c
; Sequence 25, Application US/08299849B
; Patent No. 5612201
;  GENERAL INFORMATION:
;    APPLICANT:  De Plaen, Etienne; Boon-Falleur, Thierry;
;    APPLICANT:  Leth , Bernard; Szikora, Jean-Pierre; De Smet, Charles;
;    APPLICANT:  Chomez, Patrick
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Useful In
;    TITLE OF INVENTION:  Determining Expression Of A Tumor Antigen Precursor
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/299,849B
;      FILING DATE:  1-SEPTEMBER-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/037,230
;      FILING DATE:  26-MARCH-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,364
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      APPLICATION NUMBER:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-May-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5612201man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5355
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2099 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  smage-II
US-08-299-849B-25


  Query Match             60.8%;  Score 15.2;  DB 1;  Length 2099;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       6 gccagcatttcgatcctcgt 25
```

```
          ||||||||||  |||| ||
Db    944 GCCAGCATTTCACTCCTTGT 925


RESULT  33
US-08-142-368A-25/c
; Sequence 25, Application US/08142368A
; Patent No. 5925729
;  GENERAL INFORMATION:
;    APPLICANT:  Boon-Falleur, Thierry; Van der Bruggen, Thierry;
;    APPLICANT:  Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
;    APPLICANT:  Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
;    TITLE OF INVENTION:  Tumor Rejection Antigen Precursors, Tumor
;    TITLE OF INVENTION:  Rejection Antigens and Uses Thereof
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/142,368A
;      FILING DATE:  02-MAY-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,364
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      APPLICATION NUMBER:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-May-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5925729man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5253.4-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2099 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  smage-II
US-08-142-368A-25


  Query Match             60.8%;  Score 15.2;  DB 2;  Length 2099;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      6 gccagcatttcgatcctcgt 25
          ||||||||||  |||| ||
Db    944 GCCAGCATTTCACTCCTTGT 925


RESULT  34
US-08-967-727-25/c
; Sequence 25, Application US/08967727
; Patent No. 6025474
;  GENERAL INFORMATION:
;    APPLICANT:  Gaugler, B atrice; Van den Eynde, Beno t;
;    APPLICANT:  van der Bruggen, Pierre;  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Coding For
;    TITLE OF INVENTION:  Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/967,727
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/037,230
;      FILING DATE:  26-MARCH-1993
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,365
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      FILING DATE:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 6025474man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5353
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2099 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  smage-II
US-08-967-727-25


  Query Match             60.8%;  Score 15.2;  DB 3;  Length 2099;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      6 gccagcatttcgatcctcgt 25
          ||||||||||  |||| ||
Db    944 GCCAGCATTTCACTCCTTGT 925


RESULT  35
US-08-037-230D-25/c
; Sequence 25, Application US/08037230D
```

```
; Patent No. 6235525
;  GENERAL INFORMATION:
;    APPLICANT:  Gaugler, B atrice; Van den Eynde, Beno t;
;    APPLICANT:  van der Bruggen, Pierre;  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Coding For
;    TITLE OF INVENTION:  Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/037,230D
;      FILING DATE:  26-MARCH-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,364
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      FILING DATE:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 6235525man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5353
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2099 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  smage-II
US-08-037-230D-25


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 2099;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       6 gccagcatttcgatcctcgt 25
           ||||||||||  |||| ||
Db     944 GCCAGCATTTCACTCCTTGT 925


RESULT  36
US-08-299-849B-24/c
; Sequence 24, Application US/08299849B
; Patent No. 5612201
;  GENERAL INFORMATION:
;    APPLICANT:  De Plaen, Etienne; Boon-Falleur, Thierry;
;    APPLICANT:  Leth , Bernard; Szikora, Jean-Pierre; De Smet, Charles;
;    APPLICANT:  Chomez, Patrick
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Useful In
;    TITLE OF INVENTION:  Determining Expression Of A Tumor Antigen Precursor
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/299,849B
;      FILING DATE:  1-SEPTEMBER-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/037,230
;      FILING DATE:  26-MARCH-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,364
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      APPLICATION NUMBER:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-May-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5612201man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5355
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2150 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  smage-I
US-08-299-849B-24


  Query Match             60.8%;  Score 15.2;  DB 1;  Length 2150;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       6 gccagcatttcgatcctcgt 25
           ||||||||||  |||| ||
Db     933 GCCAGCATTTCACTCCTTGT 914


RESULT  37
US-08-142-368A-24/c
; Sequence 24, Application US/08142368A
; Patent No. 5925729
;  GENERAL INFORMATION:
;    APPLICANT:  Boon-Falleur, Thierry; Van der Bruggen, Thierry;
;    APPLICANT:  Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
;    APPLICANT:  Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
;    TITLE OF INVENTION:  Tumor Rejection Antigen Precursors, Tumor
```

```
;    TITLE OF INVENTION:  Rejection Antigens and Uses Thereof
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/142,368A
;      FILING DATE:  02-MAY-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,364
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      APPLICATION NUMBER:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-May-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5925729man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5253.4-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2150 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  smage-I
US-08-142-368A-24


  Query Match             60.8%;  Score 15.2;  DB 2;  Length 2150;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       6 gccagcatttcgatcctcgt 25
           ||||||||||| |||| ||
Db     933 GCCAGCATTTCACTCCTTGT 914


RESULT  38
US-08-967-727-24/c
; Sequence 24, Application US/08967727
; Patent No. 6025474
;  GENERAL INFORMATION:
;    APPLICANT:  Gaugler, B atrice; Van den Eynde, Beno t;
;    APPLICANT:  van der Bruggen, Pierre;  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Coding For
;    TITLE OF INVENTION:  Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/967,727
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/037,230
;      FILING DATE:  26-MARCH-1993
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,365
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      FILING DATE:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 6025474man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5353
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2150 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  smage-I
US-08-967-727-24


  Query Match             60.8%;  Score 15.2;  DB 3;  Length 2150;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       6 gccagcatttcgatcctcgt 25
           ||||||||||| |||| ||
Db     933 GCCAGCATTTCACTCCTTGT 914


RESULT  39
US-08-037-230D-24/c
; Sequence 24, Application US/08037230D
; Patent No. 6235525
;  GENERAL INFORMATION:
;    APPLICANT:  Gaugler, B atrice; Van den Eynde, Beno t;
;    APPLICANT:  van der Bruggen, Pierre;  Boon-Falleur, Thierry
;    TITLE OF INVENTION:  Isolated Nucleic Acid Molecules Coding For
;    TITLE OF INVENTION:  Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      ZIP:  10022
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/037,230D
;      FILING DATE:  26-MARCH-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/04354
;      FILING DATE:  22-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/807,043
;      FILING DATE:  12-DECEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/764,364
;      FILING DATE:  23-SEPTEMBER-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/728,838
;      FILING DATE:  9-JULY-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/705,702
;      FILING DATE:  23-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 6235525man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5353
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2150 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    FEATURE:
;      NAME/KEY:  smage-I
US-08-037-230D-24


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 2150;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       6 gccagcatttcgatcctcgt 25
           ||||||||||| |||| ||
Db     933 GCCAGCATTTCACTCCTTGT 914


RESULT  40
US-09-625-188-19
; Sequence 19, Application US/09625188
; Patent No. 6307037
; GENERAL INFORMATION:
;  APPLICANT:  No. 6307037artis AG
;  TITLE OF INVENTION:  Fungal Target Genes and Methods
;  FILE REFERENCE:  PB/5-31285P1
;  CURRENT APPLICATION NUMBER: US/09/625,188
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 19
;   LENGTH: 2190
;   TYPE: DNA
;   ORGANISM: Ashbya gossypii
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(2190)
US-09-625-188-19


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 2190;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatc 20
           ||||| |||||||||  |||
Db    1362 gcatcaccagcatttgcatc 1381


RESULT  41
US-08-693-214-1/c
; Sequence 1, Application US/08693214
; Patent No. 6054318
;  GENERAL INFORMATION:
;    APPLICANT:  COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH
;    APPLICANT:  ORGANISATION
;    TITLE OF INVENTION:  Expression of the Glucose Oxidase Gene
;    TITLE OF INVENTION:  in Transgenic Organisms
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  COOPER AND DUNHAM LLP
;      STREET:  1185 AVENUE OF THE AMERICAS
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/693,214
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PM 3792/94
;      FILING DATE:  10-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  INTERNATIONAL PCT/AU95/00059
;      FILING DATE:  10-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WHITE, JOHN P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1 212 278 0400
;      TELEFAX:  1 212 391 0525
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3029 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Talaromyces flavus
;      STRAIN:  ATCC 32908
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1032..2846
US-08-693-214-1


  Query Match             60.8%;  Score 15.2;  DB 3;  Length 3029;
  Best Local Similarity   85.0%;  Pred. No. 2.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 cgccagcatttcgatcctcg 24
           ||  ||||||| ||||||||
Db    1305 CGTAAGCATTTGGATCCTCG 1286


RESULT  42
```

```
US-08-404-665-3/c
; Sequence 3, Application US/08404665
; Patent No. 5591583
;  GENERAL INFORMATION:
;    APPLICANT:  Reid, Robert A.
;    APPLICANT:  Ackley, Rhonda L.
;    APPLICANT:  Hemperly, John J.
;    TITLE OF INVENTION:  HUMAN RESTRICTIN AND NUCLEIC ACID
;    TITLE OF INVENTION:  SEQUENCES
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Rodrick, Becton Dickinson and
;      ADDRESSEE:  Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/404,665
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3341
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4724 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-404-665-3


  Query Match             60.8%;  Score 15.2;  DB 1;  Length 4724;
  Best Local Similarity   85.0%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcc 21
           | || ||||||| |||||||
Db     903 CCTCTCCAGCATCTCGATCC 884


RESULT  43
US-08-404-671-3/c
; Sequence 3, Application US/08404671
; Patent No. 5635360
;  GENERAL INFORMATION:
;    APPLICANT:  Reid, Robert A.
;    APPLICANT:  Ackley, Rhonda L.
;    APPLICANT:  Hemperly, John J.
;    TITLE OF INVENTION:  HUMAN RESTRICTIN AND NUCLEIC ACID
;    TITLE OF INVENTION:  SEQUENCES
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Rodrick, Becton Dickinson and
;      ADDRESSEE:  Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/404,671
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3341
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4724 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-404-671-3


  Query Match             60.8%;  Score 15.2;  DB 1;  Length 4724;
  Best Local Similarity   85.0%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcc 21
           | || ||||||| |||||||
Db     903 CCTCTCCAGCATCTCGATCC 884


RESULT  44
US-08-404-781-3/c
; Sequence 3, Application US/08404781
; Patent No. 5681931
;  GENERAL INFORMATION:
;    APPLICANT:  Reid, Robert A.
;    APPLICANT:  Ackley, Rhonda L.
;    APPLICANT:  Hemperly, John J.
;    TITLE OF INVENTION:  HUMAN RESTRICTIN AND NUCLEIC ACID
;    TITLE OF INVENTION:  SEQUENCES
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Rodrick, Becton Dickinson and
;      ADDRESSEE:  Company
;      STREET:  1 Becton Drive
;      CITY:  Franklin Lakes
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07417
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/404,781
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fugit, Donna R.
;      REGISTRATION NUMBER:  32,135
;      REFERENCE/DOCKET NUMBER:  P-3341
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4724 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-404-781-3


  Query Match             60.8%;  Score 15.2;  DB 1;  Length 4724;
  Best Local Similarity   85.0%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy       2 catcgccagcatttcgatcc 21
           | || ||||||| |||||||
Db     903 CCTCTCCAGCATCTCGATCC 884


RESULT  45
US-08-781-891-79/c
; Sequence 79, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  240052.419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  79:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  87350 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-781-891-79


  Query Match             60.8%;  Score 15.2;  DB 3;  Length 87350;
  Best Local Similarity   85.0%;  Pred. No. 3.1e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       6 gccagcatttcgatcctcgt 25
           |||||||||||  ||||| |
Db   33740 GCCAGCATTTCAGTCCTCAT 33721


Search completed: December 26, 2001, 12:27:04
Job time: 7381 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:11:50 ; Search time 4612.2 Seconds
(without alignments)
58.246 Million cell updates/sec

Title: US-09-396-196F-7
Perfect score: 25
Sequence: 1 gcatcgccagcatttcgatcctcgt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18.6	74.4	216	13	AZ577684	AZ577684 13b12 Sho
c	2	18.6	74.4	278	11	W10819	W10819 ma54h02.r1
c	3	18.6	74.4	538	10	AV439422	AV439422 AV439422
c	4	18.6	74.4	636	10	AW202534	AW202534 687038H06
c	5	18.4	73.6	617	10	AW637075	AW637075 bl54a09.w
c	6	18.2	72.8	307	10	BB336604	BB336604 BB336604
c	7	18.2	72.8	878	11	BF684643	BF684643 602141001
c	8	18.2	72.8	1426	11	BG851910	BG851910 1024032F1
	9	17.8	71.2	453	13	AQ068105	AQ068105 HS_2177_B
	10	17.8	71.2	472	11	BG278235	BG278235 a2f02np.r
c	11	17.8	71.2	568	10	AI519307	AI519307 LD38845.5
c	12	17.8	71.2	606	10	AA941558	AA941558 LD25751.5
	13	17.8	71.2	616	13	AZ852429	AZ852429 2M0155A15
c	14	17.8	71.2	636	10	AI108881	AI108881 GH08185.5
c	15	17.8	71.2	652	10	AI063994	AI063994 GH04236.5
c	16	17.8	71.2	654	13	TA44D09P	AL455005 T. brucei
c	17	17.6	70.4	208	10	AW203894	AW203894 SWOv3MCAM
	18	17.6	70.4	295	10	AA335698	AA335698 EST40140
c	19	17.6	70.4	353	10	AW172391	AW172391 xj38b09.x
c	20	17.6	70.4	360	10	AA994865	AA994865 ou06e08.s
c	21	17.6	70.4	400	10	AA067753	AA067753 mm56a05.r
c	22	17.6	70.4	442	10	AA294431	AA294431 SWOv3MCA8
c	23	17.6	70.4	488	13	AQ868181	AQ868181 nbeb0026J
c	24	17.6	70.4	519	10	BE372749	BE372749 601224261
c	25	17.6	70.4	529	10	AW182566	AW182566 xj44b11.x
	26	17.6	70.4	541	13	FR0002911	Z86694 F.rubripes
c	27	17.6	70.4	554	10	AW288131	AW288131 SWOv3MCAM
c	28	17.6	70.4	706	10	AW184046	AW184046 fj08a10.y
c	29	17.6	70.4	732	11	BI412264	BI412264 602987756
c	30	17.6	70.4	790	10	AL578914	AL578914 AL578914
c	31	17.6	70.4	879	11	BF161425	BF161425 601770406
	32	17.6	70.4	920	10	AL556716	AL556716 AL556716
c	33	17.6	70.4	1022	11	BF608598	BF608598 MY1_00154
c	34	17.6	70.4	1024	13	CNS06X9E	AL419448 T3 end of
c	35	17.4	69.6	281	10	AI978168	AI978168 614041A04
	36	17.4	69.6	299	10	AW355901	AW355901 707015F11
c	37	17.4	69.6	385	11	T14661	T14661 05c04a07-f2
c	38	17.4	69.6	395	10	AI834428	AI834428 606068A02
	39	17.4	69.6	471	10	AI812153	AI812153 605087H09
	40	17.4	69.6	504	10	BE511349	BE511349 946059F01
c	41	17.4	69.6	525	10	AI692089	AI692089 606011H09
c	42	17.4	69.6	528	10	AI740036	AI740036 605070A01
c	43	17.4	69.6	536	10	AI714408	AI714408 606013D04
c	44	17.4	69.6	550	10	AI834659	AI834659 606071E07
c	45	17.4	69.6	555	10	AI861139	AI861139 603012F12
c	46	17.4	69.6	565	10	AW600443	AW600443 660068C06
c	47	17.4	69.6	573	10	AI833970	AI833970 605097G09
c	48	17.4	69.6	582	10	AI668324	AI668324 605023G02
c	49	17.4	69.6	590	10	AI861291	AI861291 603018E12
c	50	17.4	69.6	592	10	AI737741	AI737741 605040F03
c	51	17.4	69.6	594	10	AI737134	AI737134 605030C04
c	52	17.4	69.6	598	11	BG265392	BG265392 1000023C0
c	53	17.4	69.6	605	10	AI668455	AI668455 605032F01
c	54	17.4	69.6	608	10	AI737657	AI737657 605036D08
c	55	17.4	69.6	608	10	AW076486	AW076486 605033A07
c	56	17.4	69.6	609	10	AI987312	AI987312 660003G08
c	57	17.4	69.6	611	10	AI861486	AI861486 614014E11
c	58	17.4	69.6	615	10	AI712151	AI712151 605065A01
c	59	17.4	69.6	616	10	AI711818	AI711818 605064D10
c	60	17.4	69.6	734	11	BG836936	BG836936 Zm08_09a0
c	61	17.2	68.8	203	11	BG589383	BG589383 EST497225
	62	17.2	68.8	294	13	AZ598170	AZ598170 1M0412O13
c	63	17.2	68.8	338	11	BE859615	BE859615 UI-M-AP0-
	64	17.2	68.8	376	13	AQ029176	AQ029176 RPCI11-41
	65	17.2	68.8	437	13	AQ988849	AQ988849 17A1F03NE
c	66	17.2	68.8	440	10	AW322034	AW322034 uo23b02.x
	67	17.2	68.8	491	10	AI399397	AI399397 NCSC3A11T
c	68	17.2	68.8	571	10	BE576144	BE576144 dc34d08.y
c	69	17.2	68.8	593	11	BG310941	BG310941 ia71c04.y
	70	17.2	68.8	594	13	AQ851575	AQ851575 CpG1410A
	71	17.2	68.8	676	10	BE036913	BE036913 MP09A10 M
	72	17.2	68.8	685	10	BE577458	BE577458 L48-2304T
c	73	17.2	68.8	686	11	BI151809	BI151809 602915976
c	74	17.2	68.8	689	10	AV675792	AV675792 AV675792
c	75	17.2	68.8	690	10	BE230792	BE230792 SWOvAMCAQ
	76	17.2	68.8	720	11	BG269350	BG269350 L0-3395T3
	77	17.2	68.8	768	10	AW266592	AW266592 L0-1310T3
c	78	17.2	68.8	907	13	CNS01NTI	AL152663 Anopheles
c	79	17.2	68.8	944	13	CNS07883	AL433657 T7 end of
	80	17.2	68.8	982	13	CNS04EUT	AL287534 Tetraodon
	81	17.2	68.8	1069	11	BF570058	BF570058 602186066
	82	17.2	68.8	1072	10	BE035550	BE035550 MO09A10 M
c	83	17.2	68.8	1090	13	CNS058GS	AL325909 Tetraodon
c	84	17.2	68.8	1101	13	CNS00EY1	AL069592 Drosophil
c	85	17.2	68.8	1330	11	BG280908	BG280908 602401316

	86	17	68.0	170	10	AV433546	AV433546 AV433546
	87	17	68.0	198	10	AI424685	AI424685 tf64c03.x
c	88	17	68.0	299	11	BF071505	BF071505 st60c12.y
	89	17	68.0	321	11	BE983721	BE983721 UI-M-CG0p
	90	17	68.0	400	11	BF478582	BF478582 WHE2003_C
c	91	17	68.0	406	13	AQ557235	AQ557235 HS_2085_A
c	92	17	68.0	412	10	AW492386	AW492386 UI-M-BH3-
c	93	17	68.0	414	10	AW370117	AW370117 MR0-BT024
c	94	17	68.0	422	10	BE057776	BE057776 sn07a08.y
	95	17	68.0	432	11	BF428707	BF428707 WHE1412_D
	96	17	68.0	441	11	BF263940	BF263940 HV_CEa000
c	97	17	68.0	483	13	AQ362407	AQ362407 nbxb0049F
	98	17	68.0	495	10	BE497164	BE497164 WHE0765_A
c	99	17	68.0	498	10	AI532390	AI532390 SD03891.5
	100	17	68.0	505	10	BE490479	BE490479 WHE0360_B

ALIGNMENTS

```
RESULT   1
AZ577684
LOCUS       AZ577684      216 bp    DNA             GSS       08-DEC-2000
DEFINITION  13b12 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
            sp. NGR234 genomic clone 13b12, DNA sequence.
ACCESSION   AZ577684
VERSION     AZ577684.1  GI:11604206
KEYWORDS    GSS.
SOURCE      Rhizobium sp. NGR234.
  ORGANISM  Rhizobium sp. NGR234
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 216)
  AUTHORS   Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.
  TITLE     Genetic snapshots of the Rhizobium species NGR234 genome
  JOURNAL   GenomeBiology.com 1 (6), 0014.1-0014.7 (2000)
COMMENT     Contact: Virginie Viprey
            Laboratoire de Biologie Moleculaire des Plantes Superieures
            University of Geneva
            1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland
            Tel: +44(0)1603450000
            Fax: +44(0)1603450045
            Email: virginie.viprey@bbsrc.ac.uk
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .216
                     /organism="Rhizobium sp. NGR234"
                     /strain="ANU265"
                     /db_xref="taxon:394"
                     /clone="13b12"
                     /clone_lib="Shot-gun genomic library of Rhizobium strain
                     ANU265"
                     /note="Vector: M13; derivative strain of NGR234 cured of
                     pNGR234a"
BASE COUNT       34 a      70 c      65 g      46 t       1 others
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 13;  Length 216;
  Best Local Similarity   84.0%;  Pred. No. 3.1e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||||||||||| |||| || | | |
Db     114 GCATCGCCAGCATGTCGAACCACAT 138


RESULT   2
W10819/c
LOCUS       W10819        278 bp    mRNA            EST       02-OCT-1997
DEFINITION  ma54h02.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:314547 5' similar to gb:J04982_cds1 ADP,ATP CARRIER PROTEIN,
            HEART/SKELETAL MUSCLE ISOFORM (HUMAN);, mRNA sequence.
ACCESSION   W10819
VERSION     W10819.1  GI:1285135
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 278)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI: 205163
            Trace considered overall poor quality
            Seq primer: mob.REGA+ET
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .278
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:314547"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
BASE COUNT       64 a      69 c      89 g      56 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 11;  Length 278;
  Best Local Similarity   84.0%;  Pred. No. 3.2e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           |||  | |||||||||| |||||||||
Db     198 GCTGCACCAGCATTTTGATCCTCGT 174


RESULT   3
AV439422/c
LOCUS       AV439422      538 bp    mRNA            EST       23-AUG-2000
DEFINITION  AV439422 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PS050c10_r 5', mRNA sequence.
ACCESSION   AV439422
VERSION     AV439422.1  GI:8594647
KEYWORDS    EST.
SOURCE      Porphyra yezoensis.
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 538)
```

```
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .538
                     /organism="Porphyra yezoensis"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PS050c10_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      114 a    158 c    170 g     96 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 10;  Length 538;
  Best Local Similarity   84.0%;  Pred. No. 3.5e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           ||||| || ||||||| |||| |||
Db     380 GCATCCCCTGCATTTCCATCCCCGT 356


RESULT   4
AW202534/c
LOCUS       AW202534      636 bp    mRNA            EST       30-MAR-2000
DEFINITION  687038H06.x1 687 - Early embryo from Delaware Zea mays cDNA, mRNA
            sequence.
ACCESSION   AW202534
VERSION     AW202534.1  GI:6501161
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 636)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 687038  row: H  column: 06.
FEATURES             Location/Qualifiers
     source          1. .636
                     /organism="Zea mays"
                     /cultivar="Illinois High Oil"
                     /db_xref="taxon:4577"
                     /clone_lib="687 - Early embryo from Delaware"
                     /tissue_type="embryo"
                     /dev_stage="14, 21, 28, and 35 days after pollination"
                     /lab_host="E. coli SOLR"
                     /note="Organ: embryo; Vector: pBluescript SK; Site_1: XhoI
                     ; Site_2: EcoRI; Library was prepared by Statagene using
                     the Uni-ZAP XR system (Stratagene BN937328-12). Clones
                     were picked by a Q-bot after blue/white selection
                     (ampicillin resistance - use 100 micrograms/microliter).
                     Developed from a pool of equal amounts of RNA from
                     developing embryos sampled at 14, 21, 28 and 35 days after
                     pollination of the Illinois High Oil Maize Strain Cycle
                     90. This closed strain has been selected for high oil
                     concentration for 90 generations and originates from the
                     1890s era open pollinated variety Burr's White"
BASE COUNT      111 a    197 c    190 g    138 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 10;  Length 636;
  Best Local Similarity   84.0%;  Pred. No. 3.6e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcgt 25
           ||| |||||||| | ||| ||||||
Db     578 GCAGCGCCAGCAATGCGAGCCTCGT 554


RESULT   5
AW637075/c
LOCUS       AW637075      617 bp    mRNA            EST       26-APR-2001
DEFINITION  bl54a09.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
            laevis cDNA clone PBX0054A09 5', mRNA sequence.
ACCESSION   AW637075
VERSION     AW637075.1  GI:7394175
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Staffa,N.G.
            Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman
            ,J.W., Bonaldo,M.F. and Soares,M.B.
  TITLE     The NIEHS Xenopus maternal EST project: interim analysis of the
            first 13,879 ESTs from unfertilized eggs
  JOURNAL   Gene 267 (1), 71-87 (2001)
  MEDLINE   21211403
COMMENT     Contact: Perry J. Blackshear
            Office of Clinical Research and Laboratory of Signal Transduction
            National Institute of Environmental Health Sciences
            A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
            USA
            Tel: 919 541-4899
            Fax: 919 541-4571
            Email: black009@niehs.nih.gov
            Clone is available through Research Genetics, Inc., 2130 Memorial
            Parkway, Huntsville, AL 35901
             phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
            cdna@resgen.com
             DNA Sequencing and analyses performed by National Institutes of
            Health Intramural Sequencing Center (NISC).
            PCR PRimers
            FORWARD: TGTAAAACGACGGCCAGT
            BACKWARD: CAGGAAACAGCTATGACC
            Plate: 0054  row: A  column: 09
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="PBX0054A09"
                     /clone_lib="Blackshear/Soares normalized Xenopus egg
                     library"
                     /sex="female"
                     /tissue_type="unfertilized egg"
                     /cell_type="unfertilized egg"
                     /dev_stage="unfertilized egg"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
                     PolyA-selected mRNA was prepared from unfertilized Xenopus
```

laevis eggs. The library was constructed in the vector pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and Soares, M.B. 'Normalization and subtraction: two approaches to facilitate gene discovery', Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dT18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT7T3-Pac vector. The library contained approximately 7.2 X 10^5 recombinants, with average insert sizes of 1-1.5 kb."
BASE COUNT 220 a 109 c 110 g 178 t
ORIGIN

Query Match 73.6%; Score 18.4; DB 10; Length 617;
Best Local Similarity 95.0%; Pred. No. 4.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 gccagcatttcgatcctcgt 25
||||||||||| ||||||||
Db 20 GCCAGCATTTCCATCCTCGT 1

RESULT 6
BB336604/c
LOCUS BB336604 307 bp mRNA EST 11-JUL-2000
DEFINITION BB336604 RIKEN full-length enriched, 10 days neonate cerebellum Mus musculus cDNA clone B930001I22 3', mRNA sequence.
ACCESSION BB336604
VERSION BB336604.1 GI:9045367
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 307)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .307
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="B930001I22"
/clone_lib="RIKEN full-length enriched, 10 days neonate cerebellum"
/sex="mixed"
/tissue_type="cerebellum"
/dev_stage="10 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
BASE COUNT 64 a 99 c 77 g 67 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 307;
Best Local Similarity 87.0%; Pred. No. 4.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gcatcgccagcatttcgatcctc 23
|||| || || ||||||||||||
Db 56 GCATGGCTAGGATTTCGATCCTC 34

RESULT 7
BF684643/c
LOCUS BF684643 878 bp mRNA EST 22-DEC-2000
DEFINITION 602141001F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4301856 5', mRNA sequence.
ACCESSION BF684643
VERSION BF684643.1 GI:11970051
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 878)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1163 row: b column: 01
High quality sequence stop: 689.
FEATURES Location/Qualifiers
source 1. .878

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4301856"
/clone_lib="NIH_MGC_46"
/tissue_type="leiomyosarcoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 250 a 202 c 306 g 120 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 878;
Best Local Similarity 87.0%; Pred. No. 5.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcgccagcatttcgatcctcgt 25
||||.||||.|||||||||.|||
Db 788 ATCGTCAGCCTTTCGATCCACGT 766

RESULT 8
BG851910/c
LOCUS BG851910 1426 bp mRNA EST 29-MAY-2001
DEFINITION 1024032F10.y2 C. reinhardtii CC-1690, normalized, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BG851910
VERSION BG851910.1 GI:14233094
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 1426)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P., McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES Location/Qualifiers
source 1. .1426
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT 314 a 347 c 349 g 388 t 28 others
ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 1426;
Best Local Similarity 80.0%; Pred. No. 5.9e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcatcgccagcatttcgatcctcgt 25
||| ||| ||||| ||| ||||||
Db 917 GCAGCGCGCGCATTNCGANCCTCGT 893

RESULT 9
AQ068105
LOCUS AQ068105 453 bp DNA GSS 04-AUG-1998
DEFINITION HS_2177_B2_E02_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2177 Col=4 Row=J, DNA sequence.
ACCESSION AQ068105
VERSION AQ068105.1 GI:3379373
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 453)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2177 row: J column: 4
Class: BAC ends
High quality sequence stop: 453.
FEATURES Location/Qualifiers
source 1. .453
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2177 Col=4 Row=J"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 117 a 108 c 73 g 152 t 3 others
ORIGIN

Query Match 71.2%; Score 17.8; DB 13; Length 453;
Best Local Similarity 90.5%; Pred. No. 7.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 tcgccagcatttcgatcctcg 24
|||||||||.||||||.||||
Db 418 TCGCCAGCACTTCGATACTCG 438

RESULT 10
BG278235
LOCUS BG278235 472 bp mRNA EST 21-FEB-2001
DEFINITION a2f02np.r1 Neurospora crassa sexual cDNA library, Uni-zap vector system Neurospora crassa cDNA clone a2f02np 5', mRNA sequence.
ACCESSION BG278235

```
VERSION     BG278235.1  GI:13074338
KEYWORDS    EST.
SOURCE      Neurospora crassa.
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariales; Sordariaceae; Neurospora.
REFERENCE   1  (bases 1 to 472)
  AUTHORS   Kupfer,D., Lai,H., Nelson,M. and Roe,B.
  TITLE     ESTS from a Neurospora crassa Sexual cDNA Library
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: a2f02np.f1
            Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
            Department of Chemistry and Biochemistry
            Advanced Center for Genome Technology, University of Oklahoma
            620 Parrington Oval, Norman, OK 73019, USA
            Tel: 405 325 4912
            Fax: 405 325 7762
            Email: broe@ou.edu
            Contact Dr. Mary Anne Nelson, Department of Biology, University of
            New Mexico, Albuquerque, NM 87131 (e-mail address manelson@unm.edu)
            regarding clone availability
            Seq primer: M13 Universal Reverse Primer
            High quality sequence stop: 421.
FEATURES             Location/Qualifiers
     source          1. .472
                     /organism="Neurospora crassa"
                     /strain="wild type"
                     /db_xref="taxon:5141"
                     /clone="a2f02np"
                     /clone_lib="Neurospora crassa sexual cDNA library, Uni-zap
                     vector system"
                     /tissue_type="perithecia (fruiting bodies)"
                     /dev_stage="sexual"
                     /lab_host="E. coli strain SOLR"
                     /note="Vector: pBlueScript SK-; Site_1: EcoRI; Site_2:
                     XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
                     ; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT       89 a    182 c    107 g     94 t
ORIGIN
```

```
  Query Match             71.2%;  Score 17.8;  DB 11;  Length 472;
  Best Local Similarity   90.5%;  Pred. No. 7.6e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 tcgccagcatttcgatcctcg 24
           |||||| || ||||||||||
Db     256 TCGCCACCACTTCGATCCTCG 276
```

```
RESULT  11
AI519307/c
LOCUS       AI519307      568 bp    mRNA            EST       19-APR-2001
DEFINITION  LD38845.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD38845 5prime, mRNA sequence.
ACCESSION   AI519307
VERSION     AI519307.1  GI:4425161
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence AC005557
            Plate: 388  row: D  column: 9
            High quality sequence stop: 422.
FEATURES             Location/Qualifiers
     source          1. .568
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD38845"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      168 a    145 c    191 g     64 t
ORIGIN
```

```
  Query Match             71.2%;  Score 17.8;  DB 10;  Length 568;
  Best Local Similarity   90.5%;  Pred. No. 7.8e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcc 21
           |||||||||||||  ||||||
Db     344 GCATCGCCAGCATCGCGATCC 324
```

```
RESULT  12
AA941558/c
LOCUS       AA941558      606 bp    mRNA            EST       19-APR-2001
DEFINITION  LD25751.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD25751 5prime, mRNA sequence.
ACCESSION   AA941558
VERSION     AA941558.1  GI:3101471
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence DS02777_4_e1
            Plate: 257  row: E  column: 3
            High quality sequence stop: 549.
FEATURES             Location/Qualifiers
     source          1. .606
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="LD25751"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      165 a    170 c    195 g     76 t
ORIGIN
```

```
  Query Match             71.2%;  Score 17.8;  DB 10;  Length 606;
```

```
  Best Local Similarity   90.5%;  Pred. No. 7.9e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcc 21
           ||||||||||||| ||||||
Db     517 GCATCGCCAGCATCGCGATCC 497


RESULT  13
AZ852429
LOCUS       AZ852429      616 bp    DNA             GSS       21-FEB-2001
DEFINITION  2M0155A15F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0155A15 F, DNA sequence.
ACCESSION   AZ852429
VERSION     AZ852429.1  GI:13039507
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0155  row: A  column: 15
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 616.
FEATURES             Location/Qualifiers
     source          1. .616
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0155A15"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      162 a    119 c    111 g    224 t
ORIGIN
```

```
  Query Match             71.2%;  Score 17.8;  DB 13;  Length 616;
  Best Local Similarity   90.5%;  Pred. No. 7.9e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatcct 22
           | ||||||||||||:|| |||||
Db     402 CCTCGCCAGCATTTCTATCCT 422


RESULT  14
AI108881/c
LOCUS       AI108881      636 bp    mRNA            EST       19-APR-2001
DEFINITION  GH08185.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH08185 5prime, mRNA sequence.
ACCESSION   AI108881
VERSION     AI108881.1  GI:3477416
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 636)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence DS02777
            Plate: 81  row: H  column: 1
            High quality sequence stop: 559.
FEATURES             Location/Qualifiers
     source          1. .636
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH08185"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
BASE COUNT      181 a    165 c    187 g    103 t
ORIGIN
```

```
  Query Match             71.2%;  Score 17.8;  DB 10;  Length 636;
  Best Local Similarity   90.5%;  Pred. No. 7.9e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcc 21
           ||||||||||||| ||||||
Db     530 GCATCGCCAGCATCGCGATCC 510


RESULT  15
AI063994/c
LOCUS       AI063994      652 bp    mRNA            EST       23-APR-2001
DEFINITION  GH04236.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH04236 5 similar to CG7971: FBan0007971
            located on: 3L 62A10-62A11;: 04/10/2001, mRNA sequence.
ACCESSION   AI063994
VERSION     AI063994.2  GI:13759486
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
```

```
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 652)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     On Jul 23, 1998 this sequence version replaced gi:3339938.
            Other_ESTs: GH04236.3prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003472: arm:3L [1513110,1816300]
            estimated-cyto:62A6-62B6: 04/10/2001
            Plate: GH.42  row: C  column: 12
            High quality sequence stop: 534
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .652
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH04236"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
BASE COUNT      187 a    167 c    191 g    107 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 10;  Length 652;
  Best Local Similarity   90.5%;  Pred. No. 8e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcc 21
          ||||||||||||  ||||||
Db    533 GCATCGCCAGCATCGCGATCC 513


RESULT  16
TA44D09P/c
LOCUS       TA44D09P      654 bp    DNA             GSS       13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 44d09, forward sequence,
            genomic survey sequence.
ACCESSION   AL455005
VERSION     AL455005.1  GI:11856458
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 654)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .654
                     /organism="Trypanosoma brucei"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="44d09"
BASE COUNT      157 a    167 c    152 g    177 t      1 others
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 13;  Length 654;
  Best Local Similarity   90.5%;  Pred. No. 8e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcc 21
          ||||||||||||  ||||||
Db    431 GCATCGCCAGCAATTCCATCC 411


RESULT  17
AW203894/c
LOCUS       AW203894      208 bp    mRNA            EST       02-DEC-1999
DEFINITION  SWOv3MCAM32A09SK Onchocerca volvulus molting L3 larva cDNA
            (SL96MLW-OvmL3) Onchocerca volvulus cDNA clone SWOv3MCAM32A09 5',
            mRNA sequence.
ACCESSION   AW203894
VERSION     AW203894.1  GI:6503412
KEYWORDS    EST.
SOURCE      Onchocerca volvulus.
  ORGANISM  Onchocerca volvulus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Onchocerca.
REFERENCE   1  (bases 1 to 208)
  AUTHORS   Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
  TITLE     Genes expressed in molting L3 larvae of Onchocerca volvulus
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            Seq primer: pBluescript SK.
FEATURES             Location/Qualifiers
     source          1. .208
                     /organism="Onchocerca volvulus"
                     /strain="Kumba, Cameroons"
                     /db_xref="taxon:6282"
                     /clone="SWOv3MCAM32A09"
                     /clone_lib="Onchocerca volvulus molting L3 larva cDNA
                     (SL96MLW-OvmL3)"
                     /dev_stage="molting L3"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Filarial nematode parasite of humans. Third-stage
                     larvae, L3,  were isolated from infected black flies in
                     Cameroon (forest strain). The L3 were cultured in 20% FCS
                     in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
                     culture. L3 of O. volvulus molt to fourth-stage larvae by
                     day 5 in culture. mRNA was isolated from approximately
                     6000 molting larvae (mL3),2000 larvae from day 1, 2 or 3
                     in culture, and converted to double-stranded cDNA using
                     reverse transcriptase and oligo(dT) followed by RNase H
                     and DNA pol I. The library was constructed in the lambda
                     Uni-Zap XR vector and has 1 x 10E6 independent
                     recombinants and the average insert size is ~1200 bp. The
```

```
                library was constructed by Sara Lustigman and Michelle
                Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
                The library is available from Dr. Sara Lustigman (email:
                slustigm@nybc.org)."
BASE COUNT       76 a     40 c     57 g     35 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 208;
  Best Local Similarity   83.3%;  Pred. No. 8.3e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcg 24
          ||||||||| |   ||||||||||
Db     56 GCATCGCCACCTCCTCGATCCTCG 33


RESULT  18
AA335698
LOCUS       AA335698      295 bp    mRNA            EST       21-APR-1997
DEFINITION  EST40140 Epididymus Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA335698
VERSION     AA335698.1  GI:1987939
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 295)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC181529
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .295
                      /organism="Homo sapiens"
                      /db_xref="ATCC (inhost):137373"
                      /db_xref="taxon:9606"
                      /clone_lib="Epididymus"
                      /sex="male"
                      /dev_stage="adult"
                      /note="Organ: epididymis; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI"
BASE COUNT       62 a     97 c     79 g     54 t      3 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 295;
  Best Local Similarity   83.3%;  Pred. No. 8.7e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catcgccagcatttcgatcctcgt 25
          |||||||| | |||| ||||||| |
Db    150 CATCGCCATCCTTTCTATCCTCCT 173


RESULT  19
AW172391/c
LOCUS       AW172391      353 bp    mRNA            EST       16-NOV-1999
DEFINITION  xj38b09.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2659481 3' similar to TR:Q12799 Q12799 T-COMPLEX PROTEIN 10A
            ;, mRNA sequence.
ACCESSION   AW172391
VERSION     AW172391.1  GI:6438339
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 345.
FEATURES              Location/Qualifiers
     source           1. .353
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2659481"
                      /clone_lib="Soares_NFL_T_GBC_S1"
                      /lab_host="DH10B"
                      /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                      a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                      Equal amounts of plasmid DNA from three normalized
                      libraries (fetal lung NbHL19W, testis NHT, and B-cell
                      NCI_CGAP_GCB1) were mixed, and ss circles were made in
                      vitro. Following HAP purification, this DNA was used as
                      tracer in a subtractive hybridization reaction. The driver
                      was PCR-amplified cDNAs from pools of 5,000 clones made
                      from the same 3 libraries. The pools consisted of
                      I.M.A.G.E. clones 297480-302087, 682632-687239,
                      726408-728711, and 729096-731399. Subtraction by Bento
                      Soares and M. Fatima Bonaldo. "
BASE COUNT       66 a     88 c    106 g     93 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 353;
  Best Local Similarity   83.3%;  Pred. No. 8.9e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcg 24
          |||||||||||| ||  || |||||
Db    210 GCATCGCCAGCTTTCTGACCCTCG 187


RESULT  20
AA994865/c
LOCUS       AA994865      360 bp    mRNA            EST       27-AUG-1998
DEFINITION  ou06e08.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1625510 3' similar to TR:Q12799 Q12799 T-COMPLEX PROTEIN 10A.
```

```
            ;contains element THR repetitive element ;, mRNA sequence.
ACCESSION   AA994865
VERSION     AA994865.1  GI:3181354
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 637   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 305.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1625510"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT       60 a      95 c     117 g      88 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 360;
  Best Local Similarity   83.3%;  Pred. No. 8.9e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcg 24
           |||||||||||  ||  ||  |||||
Db     261 GCATCGCCAGCTTTCTGACCCTCG 238


RESULT  21
AA067753/c
LOCUS       AA067753      400 bp    mRNA            EST       06-FEB-1997
DEFINITION  mm56a05.r1 Stratagene mouse embryonic carcinoma (#937317) Mus
            musculus cDNA clone IMAGE:532400 5', mRNA sequence.
ACCESSION   AA067753
VERSION     AA067753.1  GI:1566033
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:319336
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 389.
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:532400"
                     /clone_lib="Stratagene mouse embryonic carcinoma (#937317
                     )"
                     /tissue_type="carcinoma"
                     /dev_stage="embryonic"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT. P19 cell
                     line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       72 a     123 c     125 g      80 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 400;
  Best Local Similarity   83.3%;  Pred. No. 9.1e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatcgccagcatttcgatcctcg 24
           |||||||||||  ||  ||  |||||
Db     219 GCATCGCCAGCGGTCCGCTCCTCG 196


RESULT  22
AA294431/c
LOCUS       AA294431      442 bp    mRNA            EST       12-NOV-1997
DEFINITION  SWOv3MCA869SK Onchocerca volvulus molting L3 larva cDNA
            (SL96MLW-OvmL3) Onchocerca volvulus cDNA clone SWmL3CO869 5', mRNA
            sequence.
ACCESSION   AA294431
VERSION     AA294431.1  GI:2099620
KEYWORDS    EST.
SOURCE      Onchocerca volvulus.
  ORGANISM  Onchocerca volvulus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Onchocerca.
REFERENCE   1  (bases 1 to 442)
  AUTHORS   Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
  TITLE     Genes expressed in molting L3 larvae of Onchocerca volvulus
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            Seq primer: pBluescript SK.
FEATURES             Location/Qualifiers
     source          1. .442
                     /organism="Onchocerca volvulus"
                     /strain="Kumba, Cameroons"
                     /db_xref="taxon:6282"
                     /clone="SWmL3CO869"
                     /clone_lib="Onchocerca volvulus molting L3 larva cDNA
                     (SL96MLW-OvmL3)"
```

```
                     /dev_stage="molting L3"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
                     Xho I; Filarial nematode parasite of humans. Third-stage
                     larvae, L3,  were isolated from infected black flies in
                     Cameroon (forest strain). The L3 were cultured in 20% FCS
                     in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
                     culture. L3 of O. volvulus molt to fourth-stage larvae by
                     day 5 in culture. mRNA was isolated from approximately
                     6000 molting larvae (mL3),2000 larvae from day 1, 2 or 3
                     in culture, and converted to double-stranded cDNA using
                     reverse transcriptase and oligo(dT) followed by RNase H
                     and DNA pol I. The library was constructed in the lambda
                     Uni-Zap XR vector and has 1 x 10E6 independent
                     recombinants and the average insert size is ~1200 bp. The
                     library was constructed by Sara Lustigman and Michelle
                     Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
                     The library is available from Dr. Sara Lustigman (email:
                     slustigm@nybc.org)."
BASE COUNT      138 a     80 c    123 g     86 t     15 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 442;
  Best Local Similarity   83.3%;  Pred. No. 9.2e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcg 24
          ||||||||| |   ||||||||||
Db     50 GCATCGCCACCTCCTCGATCCTCG 27


RESULT  23
AQ868181/c
LOCUS       AQ868181      488 bp    DNA             GSS       03-NOV-1999
DEFINITION  nbeb0026J13r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone nbeb0026J13r, DNA sequence.
ACCESSION   AQ868181
VERSION     AQ868181.1  GI:6218736
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GGAAACAGCTATGACCATG
            Class: BAC ends
            High quality sequence start: 32
            High quality sequence stop: 328.
FEATURES             Location/Qualifiers
     source          1. .488
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbeb0026J13r"
                     /clone_lib="CUGI Rice BAC Library (EcoRI)"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBACludigo; Site_1: EcoRI; Site_2: EcoRI;
                     Rice is the most important food crop in the world. Half of
                     the world population, especially those inhabiting highly
                     populated areas of the humid tropics and subtropics, rely
                     on rice as their primary source of carbohydrate.
                     Monocotyledonous rice is a diploid plant (2n=24) with a
                     haploid genome equivalent of 431 Mbp (Arumuganathan and
                     Earle, 1991). The relatively small genome of rice, three
                     times larger than that of Arabidopsis, makes it suitable
                     for genomic studies. In order to facilitate positional
                     cloning, physical mapping and genome sequencing of rice,
                     we have constructed a BAC library from Oryza sativa,
                     Nipponbare variety using EcoRI as the cloning enzyme. The
                     library contains 55,296 clones with an average insert size
                     of 121 Kb providing approximatley 15 haploid genome
                     equivalents. The deep coverage allows the isolation a
                     particular sequence with a probability of 99.9 %. Three
                     high density filters, each containing 18,432 clones
                     (doubly spotted), represent the whole library for colony
                     screening and can be requested from the Clemson University
                     BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT       86 a    163 c    143 g     94 t      2 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 488;
  Best Local Similarity   83.3%;  Pred. No. 9.3e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catcgccagcatttcgatcctcgt 25
          |||| ||||| |||| ||||| ||
Db    104 CATCTCCAGCCTTTCCATCCTTGT 81


RESULT  24
BE372749/c
LOCUS       BE372749      519 bp    mRNA            EST       21-JUL-2000
DEFINITION  601224261F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3582806 5',
            mRNA sequence.
ACCESSION   BE372749
VERSION     BE372749.1  GI:9318112
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8738  row: i  column: 15
            High quality sequence stop: 519.
FEATURES             Location/Qualifiers
     source          1. .519
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3582806"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="3 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      104 a    145 c    158 g    110 t      2 others
```

```
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 519;
  Best Local Similarity   83.3%;  Pred. No. 9.4e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcg 24
          ||||||||||| | || ||||||
Db    206 GCATCGCCAGCGGTCCGCTCCTCG 183


RESULT  25
AW182566/c
LOCUS       AW182566      529 bp    mRNA            EST       18-NOV-1999
DEFINITION  xj44b11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2660061 3' similar to TR:Q12799 Q12799 T-COMPLEX PROTEIN 10A
            ;contains element MSR1 repetitive element ;, mRNA sequence.
ACCESSION   AW182566
VERSION     AW182566.1  GI:6451026
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 463.
FEATURES             Location/Qualifiers
     source          1. .529
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2660061"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      100 a    118 c    176 g    135 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 529;
  Best Local Similarity   83.3%;  Pred. No. 9.4e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcg 24
          |||||||||||  ||  || |||||
Db    260 GCATCGCCAGCTTTCTGACCCTCG 237


RESULT  26
FR0002911
LOCUS       FR0002911     541 bp    DNA             GSS       27-FEB-1997
DEFINITION  F.rubripes GSS sequence, clone 015A17bF6, genomic survey sequence.
ACCESSION   Z86694
VERSION     Z86694.1  GI:1883606
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes.
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y.,
            Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .541
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 015A17"
                     /clone="015A17bF6"
BASE COUNT      120 a    168 c     95 g    142 t     16 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 541;
  Best Local Similarity   83.3%;  Pred. No. 9.5e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catcgccagcatttcgatcctcgt 25
          || ||||  |||||| ||||||||
Db    297 CAGCGCCCTCATTTCCATCCTCGT 320


RESULT  27
AW288131/c
LOCUS       AW288131      554 bp    mRNA            EST       16-JAN-2000
DEFINITION  SWOv3MCAM34B09SK Onchocerca volvulus molting L3 larva cDNA
            (SL96MLW-OvmL3) Onchocerca volvulus cDNA clone SWOv3MCAM34B09 5',
            mRNA sequence.
ACCESSION   AW288131
VERSION     AW288131.1  GI:6695066
KEYWORDS    EST.
SOURCE      Onchocerca volvulus.
  ORGANISM  Onchocerca volvulus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Onchocerca.
REFERENCE   1  (bases 1 to 554)
  AUTHORS   Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
  TITLE     Genes expressed in molting L3 larvae of Onchocerca volvulus
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Steven A. Williams
            Molecular Parasitology
            Smith College Department of Biological Sciences
            Department of Biological Sciences, Clark Science Center, Smith
            College, Northampton, MA, 01063, USA
            Tel: 4135853826
            Fax: 4135853786
            Email: genome@smith.edu
            Seq primer: pBluescript SK.
FEATURES             Location/Qualifiers
     source          1. .554
                     /organism="Onchocerca volvulus"
                     /strain="Kumba, Cameroons"
                     /db_xref="taxon:6282"
                     /clone="SWOv3MCAM34B09"
                     /clone_lib="Onchocerca volvulus molting L3 larva cDNA
                     (SL96MLW-OvmL3)"
```

/dev_stage="molting L3"
/lab_host="XL1-Blue MRF'"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (mL3),2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-Zap XR vector and has 1 x 10E6 independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigm@nybc.org)."
BASE COUNT 182 a 99 c 142 g 131 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 10; Length 554;
Best Local Similarity 83.3%; Pred. No. 9.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gcatcgccagcatttcgatcctcg 24
||||||||| | ||||||||||
Db 55 GCATCGCCACCTCCTCGATCCTCG 32

RESULT 28
AW184046/c
LOCUS AW184046 706 bp mRNA EST 30-JUL-2001
DEFINITION fj08a10.y1 Zebrafish adult olfactory Danio rerio cDNA 5' similar to TR:O95071 O95071 PROGESTIN INDUCED PROTEIN. ;, mRNA sequence.
ACCESSION AW184046
VERSION AW184046.1 GI:6452560
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE 1 (bases 1 to 706)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: John Ngai cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address: www.rzpd.de)
Seq primer: T3 ET from Amersham
High quality sequence stop: 489.
FEATURES Location/Qualifiers
source 1. .706
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish adult olfactory"
/sex="mixed"
/tissue_type="Olfactory rosettes"
/dev_stage="adult"
/lab_host="D10Hb (Gibco BRL)"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This is a directionally cloned cDNA library from adult Zebrafish olfactory epithelium."
BASE COUNT 152 a 204 c 211 g 139 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 10; Length 706;
Best Local Similarity 83.3%; Pred. No. 9.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 catcgccagcatttcgatcctcgt 25
|||||||| ||| || |||||| |
Db 329 CATCGCCATCATCTCCATCCTCAT 306

RESULT 29
BI412264/c
LOCUS BI412264 732 bp mRNA EST 14-AUG-2001
DEFINITION 602987756F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5143933 5', mRNA sequence.
ACCESSION BI412264
VERSION BI412264.1 GI:15173187
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 732)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by:Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11354 row: p column: 14
High quality sequence start: 29
High quality sequence stop: 672.
FEATURES Location/Qualifiers
source 1. .732
/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:5143933"
/clone_lib="NCI_CGAP_Lu33"
/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was prepared from mRNA obtained from pooled lung tumors with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCTGTTTTTTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 184 a 185 c 197 g 165 t 1 others
ORIGIN

```
  Query Match             70.4%;  Score 17.6;  DB 11;  Length 732;
  Best Local Similarity   83.3%;  Pred. No. 9.9e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcg 24
          ||||||||||| | || ||||||
Db    216 GCATCGCCAGCGGTCCGCTCCTCG 193


RESULT  30
AL578914/c
LOCUS       AL578914      790 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL578914 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK005YE06 3
            prime, mRNA sequence.
ACCESSION   AL578914
VERSION     AL578914.1  GI:12943445
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 790)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .790
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DK005YE06"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      195 a    212 c    234 g    132 t     17 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 790;
  Best Local Similarity   83.3%;  Pred. No. 1e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catcgccagcatttcgatcctcgt 25
          |||||||| | | || ||||||||
Db    650 CATCGCCATCCTGTCTATCCTCGT 627


RESULT  31
BF161425/c
LOCUS       BF161425      879 bp    mRNA            EST       30-OCT-2000
DEFINITION  601770406F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3989880 5',
            mRNA sequence.
ACCESSION   BF161425
VERSION     BF161425.1  GI:11041578
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 879)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9200  row: k  column: 01
            High quality sequence stop: 726.
FEATURES             Location/Qualifiers
     source          1. .879
                     /organism="Mus musculus"
                     /strain="CZECH II (feral)"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3989880"
                     /clone_lib="NCI_CGAP_Lu29"
                     /tissue_type="spontaneous tumor, metastatic to mammary.
                     Stem cell origin."
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      191 a    238 c    310 g    140 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 879;
  Best Local Similarity   83.3%;  Pred. No. 1e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catcgccagcatttcgatcctcgt 25
          ||| |||||| |||||||| ||| |
Db    424 CATGGCCAGCTTTTCGATGCTCCT 401


RESULT  32
AL556716
LOCUS       AL556716      920 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL556716 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK005YE06 5
            prime, mRNA sequence.
ACCESSION   AL556716
VERSION     AL556716.1  GI:12899649
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 920)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .920
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DK005YE06"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
```

```
                Life Technologies. Contact : Feng Liang Life Technologies,
                a division of Invitrogen 9800 Medical Center Drive
                Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                Email : fliang@lifetech.com URL :
                http://fulllength.invitrogen.com"
BASE COUNT      203 a    320 c    224 g    172 t      1 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 920;
  Best Local Similarity   83.3%;  Pred. No. 1e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catcgccagcatttcgatcctcgt 25
          ||||||||  | | ||  ||||||||
Db    769 CATCGCCATCCTGTCTATCCTCGT 792


RESULT  33
BF608598/c
LOCUS       BF608598     1022 bp    mRNA            EST       01-APR-2001
DEFINITION  MY1_001540 Mouse 9-day fetus cDNA library ICRFp522 Mus musculus
            cDNA clone ICRFp522P0150 5', mRNA sequence.
ACCESSION   BF608598
VERSION     BF608598.1  GI:13505090
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1022)
  AUTHORS   Yahyawi,M., Hennig,S., Neidhardt,L., Radelof,U., Hermann,B.G.,
            Lehrach,H. and O'Brien,J.
  TITLE     Detection of a high number of novel genes in a 9-day mouse embryo
            cDNA library normalised by oligonucleotide fingerprinting
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Hennig S
            laboraty 123, dept.Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1612
            Fax: +49 30 8413 1380
            Email: hennig@molgen.mpg.de
            EST's are made from clones being representatives of clone clusters.
            Clone clusters were calculated from oligonucleotide fingerprints.
            PCR PRimers
            FORWARD: 5'-GAGCTATTCCAGAAGTAGTGA-3'
            BACKWARD: 5'-TAATACGACTCACTATAGGG-3'
            Seq primer: 5'-ATTTAGGTGACACTATAG-3'
            High quality sequence stop: 1022.
FEATURES             Location/Qualifiers
     source          1. .1022
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="ICRFp522P0150"
                     /clone_lib="Mouse 9-day fetus cDNA library ICRFp522"
                     /tissue_type="whole embryo"
                     /dev_stage="embryonic 9-day"
                     /lab_host="E.coli, XL1 blue"
                     /note="Vector: PSVSport1; Site_1: NotI; Site_2: SalI;
                     Library preparation by oligo_dT priming of RNA. Clones can
                     be ordered from the Resource Center in Berlin,
                     http://www.rzpd.de."
BASE COUNT      229 a    298 c    268 g    227 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 1022;
  Best Local Similarity   83.3%;  Pred. No. 1e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcatcgccagcatttcgatcctcg 24
          ||||||||||||  | || ||||||
Db    250 GCATCGCCAGCGGTCCGCTCCTCG 227


RESULT  34
CNS06X9E/c
LOCUS       CNS06X9E     1024 bp    DNA             GSS       06-JUL-2001
DEFINITION  T3 end of clone AX0AA039E07 of library AX0AA from strain CBS 7064
            of Pichia farinosa, genomic survey sequence.
ACCESSION   AL419448
VERSION     AL419448.1  GI:12202626
KEYWORDS    GSS.
SOURCE      Pichia farinosa.
  ORGANISM  Pichia farinosa
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 1024)
  AUTHORS   de Montigny,J., Spehner,C., Souciet,J., Tekaia,F., Dujon,B.,
            Wincker,P., Artiguenave,F. and Potier,S.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 15. Pichia
            sorbitophila
  JOURNAL   FEBS Lett. 487 (1), 87-90 (2000)
  MEDLINE   20584725
REFERENCE   2  (bases 1 to 1024)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
REFERENCE   3  (bases 1 to 1024)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1024
                     /organism="Pichia farinosa"
                     /strain="CBS 7064"
                     /db_xref="taxon:4920"
                     /clone="AX0AA039E07"
                     /clone_lib="AX0AA"
                     /note="end : T3"
BASE COUNT      312 a    232 c    194 g    284 t      2 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 1024;
  Best Local Similarity   83.3%;  Pred. No. 1e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catcgccagcatttcgatcctcgt 25
          ||||  |||  |||||||||  ||| |
Db    392 CATCTCCACCATTTCGATGCTCTT 369


RESULT  35
AI978168/c
LOCUS       AI978168     281 bp    mRNA            EST       27-AUG-1999
```

```
DEFINITION  614041A04.x2 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AI978168
VERSION     AI978168.1  GI:5791376
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 281)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614041  row: A  column: 04.
FEATURES             Location/Qualifiers
     source          1. .281
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT       74 a     79 c     67 g     60 t      1 others
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 10;  Length 281;
  Best Local Similarity   94.7%;  Pred. No. 1.1e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatc 20
           ||||||| ||||||||||||
Db     107 CATCGCCTGCATTTCGATC 89


RESULT  36
AW355901
LOCUS       AW355901       299 bp    mRNA            EST       03-FEB-2000
DEFINITION  707015F11.x1 707 - Mixed adult tissues from Walbot lab (SK) Zea
            mays cDNA, mRNA sequence.
ACCESSION   AW355901
VERSION     AW355901.1  GI:6859907
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 299)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 707015  row: F  column: 11.
FEATURES             Location/Qualifiers
     source          1. .299
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="707 - Mixed adult tissues from Walbot lab (SK
                     )"
                     /tissue_type="tassel, kernel, silk, husk, root, leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: tassel, kernel, silk, husk, root, leaf;
                     Vector: pGAD10; Site_1: EcoRI; cDNA library from fully
                     differentiated maize tissues from an active  Mutator
                     plant. Tissue ratio is 4/2/1/1/1/1 (tassel, kernel, silk,
                     husk, root, leaf).  Unidirectionally cloned."
BASE COUNT       71 a     74 c     84 g     68 t      2 others
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 10;  Length 299;
  Best Local Similarity   94.7%;  Pred. No. 1.1e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatc 20
           ||||||| ||||||||||||
Db     214 CATCGCCTGCATTTCGATC 232


RESULT  37
T14661/c
LOCUS       T14661         385 bp    mRNA            EST       17-OCT-1996
DEFINITION  05c04a07-f21 membrane-free polysomes from endosperm Zea mays cDNA
            clone 05c04a07 3' end similar to sucrose synthase, mRNA sequence.
ACCESSION   T14661
VERSION     T14661.1  GI:440640
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 385)
  AUTHORS   Shen,B., Carneiro,N., Torres-Jerez,I., Stevenson,R., McCreery,T.,
            Helentjaris,T., Baysdorfer,C., Almira,E., Ferl,R., Habben,J. and
            Larkins,B.
  TITLE     Partial sequencing and mapping of clones from two maize cDNA
            libraries
  JOURNAL   Plant Mol. Biol. 26, 1085-1101 (1994)
  MEDLINE   95111093
COMMENT     Other_ESTs: 05c04a07-t7
            Contact: The Maize cDNA Project

            Helentjaris TG (primary contact)
            Dept. of Plant Sciences
            University of Arizona
            Dept. of Plant Sciences, University of Arizona, Tucson, AZ,85721
            ph: 602-6218-746
            fax: 602-621-7186
            E-mail: helnjars@ccit.arizona.edu

            Chris Baysdorfer
            Department of Biological Sciences, School of Science
            California State University, Hayward
            Hayward, CA 94542
            ph: 510-881-3459
            fax: 510-727-2035
            E-mail: cbaysdor@s1.csuhayward.edu

            Rob Ferl
            Interdisciplinary Center for Biotechnology Research
            DNA Sequencing Core
            University of Florida
```

```
           P.O. Box 100695
           Gainesville, FL 32611-0695
           ph: 904-392-1928, ext. 301
           fax: 904-392-4072
           E-mail: robferl@nervm.nerdc.ufl.edu
           Seq primer: F21.
FEATURES             Location/Qualifiers
     source          1. .385
                     /organism="Zea mays"
                     /strain="W64A2"
                     /db_xref="taxon:4577"
                     /clone="05c04a07"
                     /clone_lib="membrane-free polysomes from endosperm"
                     /lab_host="DH10B"
                     /note="Vector: ZipLox; Site_1: SalI; Site_2: NotI; ds-cDNA
                     was prepared from oligo-dT selected mRNA by priming with a
                     NotI oligo- dT oligomer and then adding the second strand
                     to RNase-nicked DNA:RNA hybrid with DNA PolI. SalI
                     adaptors were added to the ends, the ds-cDNAs were then
                     digested with NotI and size-selected.  These were
                     directionally-cloned into the ZipLox phage vector, excised
                     as plasmids, and then individually analyzed. "
BASE COUNT      115 a    107 c     88 g     75 t
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 11;  Length 385;
  Best Local Similarity   94.7%;  Pred. No. 1.1e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatc 20
           ||||||| ||||||||||
Db     184 CATCGCCTGCATTTCGATC 166


RESULT  38
AI834428/c
LOCUS       AI834428      395 bp    mRNA            EST       02-FEB-2000
DEFINITION  606068A02.x1 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI834428
VERSION     AI834428.1  GI:5468637
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606068  row: A  column: 02.
FEATURES             Location/Qualifiers
     source          1. .395
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="606 - Ear tissue cDNA library from Schmidt
                     lab"
                     /tissue_type="mixed"
                     /dev_stage="ear length from 0.5 cm - 2.0 cm"
                     /lab_host="XLOLR (Stratagene)"
                     /note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
                     ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                     lab"
BASE COUNT      115 a     99 c     94 g     87 t
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 10;  Length 395;
  Best Local Similarity   94.7%;  Pred. No. 1.1e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatc 20
           ||||||| ||||||||||
Db     229 CATCGCCTGCATTTCGATC 211


RESULT  39
AI812153
LOCUS       AI812153      471 bp    mRNA            EST       02-FEB-2000
DEFINITION  605087H09.y1 605 - Endosperm cDNA library from Schmidt lab Zea mays
            cDNA, mRNA sequence.
ACCESSION   AI812153
VERSION     AI812153.1  GI:5398809
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 605087  row: H  column: 09.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="605 - Endosperm cDNA library from Schmidt lab"
                     /tissue_type="nucellar, embryo, and endosperm"
                     /dev_stage="10-14 days post-pollination"
                     /lab_host="DH5(alpha)"
                     /note="Organ: Kernel; Vector: pAD-GAL4-2'; Site_1: EcoRI;
                     Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
                     lab"
BASE COUNT      124 a    106 c    116 g    125 t
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 10;  Length 471;
  Best Local Similarity   94.7%;  Pred. No. 1.1e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatc 20
           ||||||| ||||||||||
Db     203 CATCGCCTGCATTTCGATC 221


RESULT  40
BE511349
LOCUS       BE511349      504 bp    mRNA            EST       07-AUG-2000
DEFINITION  946059F01.y1 946 - tassel primordium prepared by Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   BE511349
VERSION     BE511349.1  GI:9732597
KEYWORDS    EST.
```

```
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 946059  row: F  column: 01.
FEATURES             Location/Qualifiers
     source          1. .504
                     /organism="Zea mays"
                     /cultivar="OH43"
                     /db_xref="taxon:4577"
                     /clone_lib="946 - tassel primordium prepared by Schmidt
                     lab"
                     /tissue_type="tassels"
                     /dev_stage="just after the transition from vegetative to
                     inflorescence development"
                     /lab_host="XLOLR"
                     /note="Organ: tassels; Vector: HybriZAP; Site_1: EcoRI;
                     Site_2: XhoI; George Chuck dissected immature tassels
                     between 1mm and 3mm.  Sharon  Stanfield prepared the cDNA
                     library in HybriZAP.  Sample insert size  range was 350 bp
                     to 3 Kb with a 1 Kb average."
BASE COUNT      122 a    113 c    133 g    136 t
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 10;  Length 504;
  Best Local Similarity   94.7%;  Pred. No. 1.1e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatc 20
           ||||||| ||||||||||||
Db     276 CATCGCCTGCATTTCGATC 294


RESULT  41
AI692089/c
LOCUS       AI692089      525 bp    mRNA            EST       02-FEB-2000
DEFINITION  606011H09.x1 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI692089
VERSION     AI692089.1  GI:4967187
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606011  row: H  column: 09.
FEATURES             Location/Qualifiers
     source          1. .525
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="606 - Ear tissue cDNA library from Schmidt
                     lab"
                     /tissue_type="mixed"
                     /dev_stage="ear length from 0.5 cm - 2.0 cm"
                     /lab_host="XLOLR (Stratagene)"
                     /note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
                     ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                     lab"
BASE COUNT      142 a    137 c    124 g    121 t      1 others
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 10;  Length 525;
  Best Local Similarity   94.7%;  Pred. No. 1.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatc 20
           ||||||| ||||||||||||
Db     231 CATCGCCTGCATTTCGATC 213


RESULT  42
AI740036/c
LOCUS       AI740036      528 bp    mRNA            EST       02-FEB-2000
DEFINITION  605070A01.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
            cDNA, mRNA sequence.
ACCESSION   AI740036
VERSION     AI740036.1  GI:5108323
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 605070  row: A  column: 01.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="605 - Endosperm cDNA library from Schmidt lab"
                     /tissue_type="nucellar, embryo, and endosperm"
                     /dev_stage="10-14 days post-pollination"
                     /lab_host="DH5(alpha)"
                     /note="Organ: Kernel; Vector: pAD-GAL4-2'; Site_1: EcoRI;
                     Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
                     lab"
BASE COUNT      137 a    148 c    133 g    109 t      1 others
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 10;  Length 528;
  Best Local Similarity   94.7%;  Pred. No. 1.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 catcgccagcatttcgatc 20
```

```
          ||||||| ||||||||||
Db    144 CATCGCCTGCATTTCGATC 126


RESULT  43
AI714408/c
LOCUS       AI714408      536 bp    mRNA            EST       02-FEB-2000
DEFINITION  606013D04.x2 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI714408
VERSION     AI714408.1  GI:5018215
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606013  row: D  column: 04.
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="606 - Ear tissue cDNA library from Schmidt
                     lab"
                     /tissue_type="mixed"
                     /dev_stage="ear length from 0.5 cm - 2.0 cm"
                     /lab_host="XLOLR (Stratagene)"
                     /note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
                     ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                     lab"
BASE COUNT      144 a    142 c    125 g    124 t      1 others
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 10;  Length 536;
  Best Local Similarity   94.7%;  Pred. No. 1.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 catcgccagcatttcgatc 20
          ||||||| ||||||||||
Db    239 CATCGCCTGCATTTCGATC 221


RESULT  44
AI834659/c
LOCUS       AI834659      550 bp    mRNA            EST       02-FEB-2000
DEFINITION  606071E07.x1 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI834659
VERSION     AI834659.1  GI:5468868
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606071  row: E  column: 07.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="606 - Ear tissue cDNA library from Schmidt
                     lab"
                     /tissue_type="mixed"
                     /dev_stage="ear length from 0.5 cm - 2.0 cm"
                     /lab_host="XLOLR (Stratagene)"
                     /note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
                     ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                     lab"
BASE COUNT      145 a    143 c    132 g    130 t
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 10;  Length 550;
  Best Local Similarity   94.7%;  Pred. No. 1.2e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 catcgccagcatttcgatc 20
          ||||||| ||||||||||
Db    232 CATCGCCTGCATTTCGATC 214


RESULT  45
AI861139/c
LOCUS       AI861139      555 bp    mRNA            EST       19-JUL-1999
DEFINITION  603012F12.x1 603 - stressed root cDNA library from Wang/Bohnert lab
            Zea mays cDNA, mRNA sequence.
ACCESSION   AI861139
VERSION     AI861139.1  GI:5525300
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 603012  row: F  column: 12.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone_lib="603 - stressed root cDNA library from
                     Wang/Bohnert lab"
                     /tissue_type="seedling"
                     /dev_stage="salt stress"
                     /lab_host="E. coli XL Gold"
                     /note="Organ: root; Vector: pBluescriptII SK(+) XR;
                     Seedling stressed root cDNA library from Wang/Bohnert lab"
```

BASE COUNT 140 a 148 c 137 g 130 t
ORIGIN

Query Match 69.6%; Score 17.4; DB 10; Length 555;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       2 catcgccagcatttcgatc 20
           ||||||| ||||||||||
Db     214 CATCGCCTGCATTTCGATC 196
```

Search completed: December 26, 2001, 12:11:59
Job time: 6481 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:58:04 ; Search time 1755.64 Seconds
(without alignments)
234.917 Million cell updates/sec

Title: US-09-396-196F-8
Perfect score: 25
Sequence: 1 gcatttcgatcctcgtcaggtgcag 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result No.		Score	Query Match	Length	DB	ID	Description
1		25	100.0	1041	6	AR029499	AR029499 Sequence
2		25	100.0	1041	6	AR034916	AR034916 Sequence
3		25	100.0	1084	6	A11530	A11530 BioB gene o
4		25	100.0	1121	6	E00893	E00893 Genomic DNA
5		25	100.0	5793	1	ECOBIO	J04423 E.coli 7,8-
6		25	100.0	5872	6	A38246	A38246 Sequence 1
7		25	100.0	5872	6	A38251	A38251 Sequence 6
8		25	100.0	5872	6	A93674	A93674 Sequence 1
9		25	100.0	5872	6	A93679	A93679 Sequence 6
10		25	100.0	5872	6	AR101809	AR101809 Sequence
11		25	100.0	5872	6	AR101810	AR101810 Sequence
12		25	100.0	11022	1	AE000180	AE000180 Escherich
13		25	100.0	13501	1	AE005258	AE005258 Escherich
14		25	100.0	297816	1	AP002553	AP002553 Escherich
15		23.4	93.6	5526	1	AF250776	AF250776 Unculture
16		21.8	87.2	4000	1	AF250768	AF250768 Unculture
17		21	84.0	128	1	ECOBIOB	M27731 Escherichia
18		20.2	80.8	7215	1	SMABIO	D17468 S. marcesce
19		18.6	74.4	1041	1	EHBIOTOP2	U38648 Erwinia her
20	c	18.6	74.4	101721	9	AC067950	AC067950 Homo sapi
21		18.6	74.4	121524	2	AF165146	AF165146 Homo sapi
22		18.6	74.4	176494	2	AC021450	AC021450 Homo sapi
23		18.4	73.6	118100	2	AC017911	AC017911 Drosophil
24	c	18.4	73.6	171705	3	AC007476	AC007476 Drosophil
25	c	18.4	73.6	266133	3	AE003822	AE003822 Drosophil
26	c	18.2	72.8	154270	2	AC023391	AC023391 Homo sapi
27		17.8	71.2	1946	1	AB061540	AB061540 Desulfovi
28	c	17.8	71.2	144483	2	AC091701	AC091701 Trypanoso
29	c	17.6	70.4	51	6	AX157421	AX157421 Sequence
30	c	17.6	70.4	51	6	AX157422	AX157422 Sequence
31		17.6	70.4	295	6	AX185944	AX185944 Sequence
32		17.6	70.4	321	6	AX185204	AX185204 Sequence
33		17.6	70.4	1315	9	AK024633	AK024633 Homo sapi
34	c	17.6	70.4	10728	9	AF298117	AF298117 Homo sapi
35	c	17.6	70.4	38340	2	F6E21	AL049914 Arabidops
36		17.6	70.4	72968	9	AC016941	AC016941 Homo sapi
37		17.6	70.4	86677	2	AC005646	AC005646 Drosophil
38		17.6	70.4	110000	2	LMFLCHR15_6	Continuation (7 of
39		17.6	70.4	143805	9	HSDJ335E1	AL109920 Human DNA
40		17.6	70.4	153064	2	AC090711	AC090711 Homo sapi
41	c	17.6	70.4	157254	2	AC023079	AC023079 Homo sapi
42		17.6	70.4	160903	3	AC007580	AC007580 Drosophil
43		17.6	70.4	169689	3	AC008306	AC008306 Drosophil
44		17.6	70.4	171398	2	AL160251	AL160251 Homo sapi
45		17.6	70.4	174424	9	HSU52112	U52112 Homo sapien
46		17.6	70.4	184043	2	AC090296	AC090296 Homo sapi
47	c	17.6	70.4	188789	2	AP001446	AP001446 Homo sapi
48	c	17.6	70.4	194920	8	ATCHRIV74	AL161578 Arabidops
49	c	17.6	70.4	202733	2	AC023200	AC023200 Homo sapi
50	c	17.6	70.4	208101	9	CNS01RI2	AL162831 Human chr
51	c	17.6	70.4	242832	9	CNS01RHK	AL161757 Human chr
52		17.6	70.4	260713	3	AE003814	AE003814 Drosophil
53		17.4	69.6	59999	9	AL499627	AL499627 Human DNA
54	c	17.2	68.8	570	8	AF236665	AF236665 Aspergill
55		17.2	68.8	876	6	AR014297	AR014297 Sequence
56		17.2	68.8	876	10	RNGLOA2U	X13295 Rat mRNA fo
57	c	17.2	68.8	3065	3	AF257606	AF257606 Drosophil
58	c	17.2	68.8	3068	3	AF257598	AF257598 Drosophil
59	c	17.2	68.8	3068	3	AF257604	AF257604 Drosophil
60	c	17.2	68.8	3078	3	AF257602	AF257602 Drosophil
61	c	17.2	68.8	3085	3	AF257607	AF257607 Drosophil
62	c	17.2	68.8	3085	3	AF257608	AF257608 Drosophil
63	c	17.2	68.8	4509	3	DME271350	AJ271350 Drosophil
64	c	17.2	68.8	16663	1	AE007074	AE007074 Mycobacte
65	c	17.2	68.8	38110	1	MTCY427	Z70692 Mycobacteri
66		17.2	68.8	42503	9	HSU115G11	Z71187 Human DNA s
67		17.2	68.8	91040	2	AC017292	AC017292 Drosophil
68	c	17.2	68.8	112366	9	AL591804	AL591804 Human DNA
69	c	17.2	68.8	161374	2	AC037430	AC037430 Homo sapi
70	c	17.2	68.8	166543	3	AC005813	AC005813 Drosophil

c	71	17.2	68.8	179947	2	AL591853	AL591853 Homo sapi
c	72	17.2	68.8	183370	3	AC008213	AC008213 Drosophil
	73	17.2	68.8	186008	2	AC018776	AC018776 Homo sapi
	74	17.2	68.8	195886	2	AC073030	AC073030 Homo sapi
	75	17.2	68.8	201938	2	AC012325	AC012325 Homo sapi
	76	17.2	68.8	216021	9	HUAC004787	AC004787 Homo sapi
c	77	17.2	68.8	256242	3	AE003756	AE003756 Drosophil
c	78	17.2	68.8	299350	1	SME591786	AL591786 Sinorhizo
	79	17.2	68.8	300950	1	AP001516	AP001516 Bacillus
	80	17	68.0	965	1	AF250770	AF250770 Unculture
c	81	17	68.0	1708	14	AF207594	AF207594 Apium vir
c	82	17	68.0	2240	8	ATU28422	U28422 Arabidopsis
c	83	17	68.0	3001	10	AF115383	AF115383 Mus muscu
	84	17	68.0	3076	8	AF037226	AF037226 Neurospor
c	85	17	68.0	3234	6	E03086	E03086 cDNA encodi
	86	17	68.0	5434	8	AF037227	AF037227 Neurospor
c	87	17	68.0	7931	10	AF396877	AF396877 Mus muscu
c	88	17	68.0	10471	3	AF008228	AF008228 Drosophil
	89	17	68.0	10562	1	AE004870	AE004870 Pseudomon
	90	17	68.0	10593	1	AE003860	AE003860 Xylella f
c	91	17	68.0	10619	1	AE000208	AE000208 Escherich
c	92	17	68.0	11451	3	DMTENM	X73154 D.melanogas
c	93	17	68.0	11922	1	AE004547	AE004547 Pseudomon
c	94	17	68.0	12445	3	DROZBPA	M31617 D.melanogas
	95	17	68.0	12736	1	AE006042	AE006042 Pasteurel
c	96	17	68.0	14045	3	DMODZ4	AF008227 Drosophil
c	97	17	68.0	14546	3	DMTTHORAX	Z31725 D.melanogas
	98	17	68.0	14698	1	AE007146	AE007146 Mycobacte
c	99	17	68.0	19418	1	D90743	D90743 Escherichia
	100	17	68.0	23471	1	AY036109	AY036109 Xanthomon

ALIGNMENTS

```
RESULT   1
AR029499
LOCUS       AR029499      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1  GI:5941472
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.Andrew.
  TITLE     Enhanced biotin biosynthesis in plant tissue
  JOURNAL   Patent: US 5859335-A 7 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.085;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||||||||||||||||||||||||
Db     99 GCATTTCGATCCTCGTCAGGTGCAG 123


RESULT   2
AR034916
LOCUS       AR034916      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1  GI:5950521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.A.
  TITLE     Transgenic plants having increased biotin content
  JOURNAL   Patent: US 5869719-A 7 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.085;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||||||||||||||||||||||||
Db     99 GCATTTCGATCCTCGTCAGGTGCAG 123


RESULT   3
A11530
LOCUS       A11530        1084 bp    DNA             PAT       09-FEB-1994
DEFINITION  BioB gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1  GI:490218
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   .
  JOURNAL   Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            24. .1064
                     /gene="BIOB"
     CDS             24. .1064
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490219"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
BASE COUNT      271 a    286 c    318 g    209 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 0.085;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||||||||||||||||||||||||
Db    122 GCATTTCGATCCTCGTCAGGTGCAG 146


RESULT   4
E00893
LOCUS       E00893        1121 bp    DNA             PAT       29-SEP-1997
DEFINITION  Genomic DNA encoding biotin Synthetase.
```

```
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES              Location/Qualifiers
     source           1. .1121
                      /organism="Escherichia coli"
                      /db_xref="taxon:562"
BASE COUNT      289 a    296 c    325 g    211 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 0.085;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||||||||||||||||||||||||
Db    140 GCATTTCGATCCTCGTCAGGTGCAG 164


RESULT   5
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES              Location/Qualifiers
     source           1. .5793
                      /organism="Escherichia coli"
                      /strain="K-12"
                      /db_xref="taxon:562"
     CDS              complement(98. .574)
                      /note="ORF 1"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="AAA23513.1"
                      /db_xref="GI:455168"
                      /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                      SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                      YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene             complement(633. .1925)
                      /gene="bioA"
     CDS              complement(633. .1925)
                      /gene="bioA"
                      /codon_start=1
                      /transl_table=11
                      /product="7,8-diamino-pelargonic acid aminotransferase"
                      /protein_id="AAA23514.1"
                      /db_xref="GI:457106"
                      /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                      LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                      ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                      SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                      GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                      LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                      QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                      PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene             2012. .3052
                      /gene="bioB"
     CDS              2012. .3052
                      /gene="bioB"
                      /codon_start=1
                      /transl_table=11
                      /product="biotin synthetase"
                      /protein_id="AAA23515.1"
                      /db_xref="GI:145425"
                      /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                      LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                      KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                      YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                      VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                      FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                      PDTDEYYNAAAL"
     gene             3049. .4203
                      /gene="bioF"
     CDS              3049. .4203
                      /gene="bioF"
                      /codon_start=1
                      /transl_table=11
                      /product="7-keto-8-amino pelargonic acid synthetase"
                      /protein_id="AAA23516.1"
                      /db_xref="GI:145426"
                      /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                      SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                      LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                      LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                      SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                      RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                      AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene             4190. .4945
                      /gene="bioC"
     CDS              4190. .4945
                      /gene="bioC"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="AAA23517.1"
                      /db_xref="GI:145427"
                      /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                      HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                      TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
```

```
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4938. .5597
                     /gene="bioD"
     CDS             4938. .5597
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAA23518.1"
                     /db_xref="GI:145428"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAC
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                     T"
BASE COUNT     1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match           100.0%;  Score 25;  DB 1;  Length 5793;
  Best Local Similarity  100.0%;  Pred. No. 0.091;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||||||||||||||||||||||||
Db   2110 GCATTTCGATCCTCGTCAGGTGCAG 2134


RESULT   6
A38246
LOCUS       A38246       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /number=1
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAA02325.1"
                     /db_xref="GI:2294845"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=3
                     /product="PROTEIN"
                     /protein_id="CAA02326.1"
                     /db_xref="GI:2294846"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="B1OA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:
                     8-AMINO-7-OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /number=5
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAA02327.1"
                     /db_xref="GI:2294847"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=6
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAA02328.1"
                     /db_xref="GI:2294848"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
```

```
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.091;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||||||||||||||||||||||||
Db    215 GCATTTCGATCCTCGTCAGGTGCAG 239


RESULT   7
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.091;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||||||||||||||||||||||||
Db    215 GCATTTCGATCCTCGTCAGGTGCAG 239


RESULT   8
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
```

```
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN
```

```
  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.091;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||||||||||||||||||||||||
Db     215 GCATTTCGATCCTCGTCAGGTGCAG 239
```

```
RESULT   9
A93679
LOCUS       A93679       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN
```

```
  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.091;
```

```
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||||||||||||||||||||||||
Db    215 GCATTTCGATCCTCGTCAGGTGCAG 239


RESULT  10
AR101809
LOCUS       AR101809     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 1 from patent US 6083712.
ACCESSION   AR101809
VERSION     AR101809.1  GI:12812607
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.091;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||||||||||||||||||||||||
Db    215 GCATTTCGATCCTCGTCAGGTGCAG 239


RESULT  11
AR101810
LOCUS       AR101810     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 6 from patent US 6083712.
ACCESSION   AR101810
VERSION     AR101810.1  GI:12812608
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.091;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||||||||||||||||||||||||
Db    215 GCATTTCGATCCTCGTCAGGTGCAG 239


RESULT  12
AE000180
LOCUS       AE000180    11022 bp    DNA             BCT       01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 70 of 400 of the complete
            genome.
ACCESSION   AE000180 U00096
VERSION     AE000180.1  GI:1786988
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 11022)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 11022)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .11022
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   <1. .130
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 4 REP sequences"
     gene            complement(147. .1430)
                     /gene="ybhC"
                     /note="b0772"
```

```
CDS             complement(147. .1430)
                /gene="ybhC"
                /function="orf; Not classified"
                /note="f427; 98 pct identical to fragment YBHC_ECOLI
                SW:P46130 (300 aa) but contains 127 additional C-terminal
                residues"
                /codon_start=1
                /transl_table=11
                /product="putative pectinesterase"
                /protein_id="AAC73859.1"
                /db_xref="GI:1786989"
                /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPAGTQGVTHTTIQAAVD
                AAIIKRTNKRQYIAVMPGEYQGTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
                NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSN1YYGFLAVNSRFN
                AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
promoter        complement(1507. .1535)
                /note="factor Sigma70; predicted +1 start at 806574"
gene            complement(1582. .2058)
                /gene="ybhB"
                /note="b0773"
CDS             complement(1582. .2058)
                /gene="ybhB"
                /function="orf; Unknown"
                /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC73860.1"
                /db_xref="GI:1786990"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
protein_bind    complement(2108. .2124)
                /note="central position to predicted promoter:85.5"
                /bound_moiety="RhaS predicted site"
gene            complement(2117. .3406)
                /gene="bioA"
                /note="b0774"
CDS             complement(2117. .3406)
                /gene="bioA"
                /EC_number="2.6.1.62"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                /codon_start=1
                /transl_table=11
                /product="7,8-diaminopelargonic acid synthetase"
                /protein_id="AAC73861.1"
                /db_xref="GI:1786991"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
promoter        complement(2193. .2221)
                /gene="bioA"
                /note="factor Sigma70; predicted +1 start at 807260"
protein_bind    3411. .3450
                /note="central position to bioB promoter: -20"
                /bound_moiety="BioB documented site"
protein_bind    3411. .3450
                /note="central position to predicted promoter: -20;genetic
                evidence for the site"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to predicted promoter:50"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to bioA promoter:50"
                /bound_moiety="BioB documented site"
promoter        3413. .3441
                /note="factor Sigma70; promoter bioB; documented +1 at
                808525"
promoter        complement(3447. .3473)
                /note="factor Sigma70; promoter bioA; documented +1 at
                808515"
gene            3493. .4533
                /gene="bioB"
                /note="b0775"
CDS             3493. .4533
                /gene="bioB"
                /EC_number="2.8.1.-"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                /codon_start=1
                /transl_table=11
                /product="biotin synthesis, sulfur insertion?"
                /protein_id="AAC73862.1"
                /db_xref="GI:1786992"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            4530. .5684
                /gene="bioF"
                /note="b0776"
CDS             4530. .5684
                /gene="bioF"
                /EC_number="2.3.1.47"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                /codon_start=1
                /transl_table=11
                /product="8-amino-7-oxononanoate synthase"
                /protein_id="AAC73863.1"
                /db_xref="GI:1786993"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
```

```
  Query Match             100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity   100.0%;  Pred. No. 0.094;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||||||||||||||||||||||||
Db    3591 GCATTTCGATCCTCGTCAGGTGCAG 3615


RESULT  13
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
```

```
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .13501
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .7576
                     /note="O-island #36; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655; Cryptic prophage
                     CP-933K; includes one copy of the 13 bp direct repeat that
                     flanks the prophage"
     gene            66. .665
                     /gene="lomK"
                     /note="Z0981"
     CDS             66. .665
                     /gene="lomK"
                     /function="putative membrane; Other or unknown (Phage or
                     Prophage Related)"
                     /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                     residues 1 to 199 of 199 from GenPept 118 :
                     gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
                     [prophage P-EibA]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein of prophage
                     CP-933K"
                     /protein_id="AAG55137.1"
                     /db_xref="GI:12513752"
                     /translation="MRKVCAAILSAAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG
                     SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG
                     PSVRVNEWFSAYAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW
                     GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"
     gene            719. .2041
                     /gene="Z0982"
     CDS             719. .2041
                     /gene="Z0982"
                     /function="putative structure; Structural component (Phage
                     or Prophage Related)"
                     /note="Residues 164 to 440 of 440 are 68.79 pct identical
                     to residues 381 to 645 of 645 from GenPept 118 :
                     gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
                     tail fiber protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tail component of prophage CP-933K"
                     /protein_id="AAG55138.1"
                     /db_xref="GI:12513753"
                     /translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE
                     NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQPGTLNDFLGAMTEDDAR
                     PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE
                     SARQAAESAASAKKSEEASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA
                     TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE
                     RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA
                     GPQGPKGETGAAGPVGATGPQGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI
                     TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"
     gene            2004. .2312
                     /gene="Z0984"
     CDS             2004. .2312
                     /gene="Z0984"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 102 of 102 are 98.03 pct identical to
                     residues 28 to 129 of 129 from GenPept 118 :
                     gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
                     hypothetical protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55139.1"
                     /db_xref="GI:12513754"
                     /translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR
                     LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK
                     "
     gene            2489. .3469
                     /gene="Z0985"
     CDS             2489. .3469
                     /gene="Z0985"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 13 to 326 of 326 are 55.73 pct identical
                     to residues 18 to 331 of 336 from GenPept 118 :
                     gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
                     predicted by Glimmer [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55140.1"
                     /db_xref="GI:12513755"
                     /translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF
                     QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS
                     KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR
                     EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD
                     MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD
                     AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"
     gene            3530. .4522
                     /gene="Z0986"
     CDS             3530. .4522
                     /gene="Z0986"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55141.1"
                     /db_xref="GI:12513756"
                     /translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
                     SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK
                     VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY
                     VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG
                     WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT
                     EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"
     gene            5350. .6231
                     /gene="Z0989"
     CDS             5350. .6231
                     /gene="Z0989"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55142.1"
                     /db_xref="GI:12513757"
                     /translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV
```

```
                     KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGEISRKIEVMLNASPVSAAM
                     ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                     FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDIIGIRMDKINGESLLNISSLPAQ
                     AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                     SYHGGKQDLISVVLSKI"
     gene            6462. .7160
                     /gene="Z0990"
     CDS             6462. .7160
                     /gene="Z0990"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55143.1"
                     /db_xref="GI:12513758"
                     /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLRF
                     QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                     IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                     SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                     VSLGFQQVRLHPLL"
     gene            complement(7668. .8144)
                     /gene="ybhB"
                     /note="Z0992"
     CDS             complement(7668. .8144)
                     /gene="ybhB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                     residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                     MG1655: B0773"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55144.1"
                     /db_xref="GI:12513759"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(8203. .9492)
                     /gene="bioA"
                     /note="Z0993"
     CDS             complement(8203. .9492)

  Query Match             100.0%;  Score 25;  DB 1;  Length 13501;
  Best Local Similarity   100.0%;  Pred. No. 0.095;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||||||||||||||||||||||||
Db    9677 GCATTTCGATCCTCGTCAGGTGCAG 9701


RESULT  14
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .1245
                     /gene="ECs0753"
     CDS             79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
     gene            1245. .2114
                     /gene="ECs0754"
     CDS             1245. .2114
                     /gene="ECs0754"
                     /note="similar to SUCD_ECOLI gi|1786949 percent identity
                     100 in 289 aa (Conserved in E.coli K-12)"
                     /codon_start=1
```

```
                /transl_table=11
                /evidence=not_experimental
                /product="succinyl-CoA synthetase alpha subunit"
                /protein_id="BAB34177.1"
                /db_xref="GI:13360213"
                /translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
                GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
                LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
                AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
                AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
                LADIGEALKTVLK"
gene            2781. .3686
                /gene="ECs0755"
CDS             2781. .3686
                /gene="ECs0755"
                /note="probable transcriptional regulator,similar to
                transcriptional regulators e.g. glycine cleavage system
                transcription activator (gcv operon activator) -
                Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
                identity 31 in 300 aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transcription regulator"
                /protein_id="BAB34178.1"
                /db_xref="GI:13360214"
                /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
                EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
                PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
                LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
                NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
                EKSDSPHLQQFREWLLAKSVVPQEM"
gene            complement(3720. .4322)
                /gene="ECs0756"
CDS             complement(3720. .4322)
                /gene="ECs0756"
                /note="probable cob(I)alamin adenosyltransferase,similar
                to cob(I)alamin adenosyltransferases (corrinoid
                adenosyltransferases) e.g. [Escherichia coli]
                gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative cob(I)alamin adenosyltransferase"
                /protein_id="BAB34179.1"
                /db_xref="GI:13360215"
                /translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
                KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
                QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
                IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"
gene            complement(4332. .5984)
                /gene="ECs0757"
CDS             complement(4332. .5984)
                /gene="ECs0757"
                /note="probable fumarate hydratase,similar to fumarate
                hydratases e.g. fumarate hydratase class I, aerobic
                (fumarase) - Escherichia coli
                gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative fumarate hydratase"
                /protein_id="BAB34180.1"
                /db_xref="GI:13360216"
                /translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
                ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
                NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
                TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
                GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
                EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
                EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
                DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
                QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSIGGAAALLAQEYVKSLRCLEYPELG
                MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"
gene            complement(6092. .7372)
                /gene="ECs0758"
CDS             complement(6092. .7372)
                /gene="ECs0758"
                /note="probable transport protein,similar to
                glutamate/aspartate transport proteins (proton glutamate
                sympport proteins) e.g. [Bacillus stearothermophilus]
                gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
                aa,also similar to C4-dicarboxylate transpor"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transport protein"
                /protein_id="BAB34181.1"
                /db_xref="GI:13360217"
                /translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
                TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
                NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
                FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
                VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
                GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
                LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
                AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"
gene            complement(7533. .7850)
                /gene="ECs0759"
CDS             complement(7533. .7850)
                /gene="ECs0759"
                /note="unknown"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="hypothetical protein"
                /protein_id="BAB34182.1"

  Query Match             100.0%;  Score 25;  DB 1;  Length 297816;
  Best Local Similarity   100.0%;  Pred. No. 0.11;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||||||||||||||||||||||||
Db  92409 GCATTTCGATCCTCGTCAGGTGCAG 92433


RESULT  15
AF250776
LOCUS       AF250776     5526 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
            modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
            biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
            biotin biosynthesis protein BioC (bioC) genes, complete cds; and
            dethiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION   AF250776
VERSION     AF250776.1  GI:12620124
KEYWORDS    .
SOURCE      uncultured bacterium pCosHE2.
  ORGANISM  uncultured bacterium pCosHE2
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
```

```
source          1. .5526
                /organism="uncultured bacterium pCosHE2"
                /db_xref="taxon:143797"
                /clone="pCosHE2"
                /note="unknown organism, cosmid clone derived from
                environmental consortium"
CDS             complement(52. .528)
                /note="ORF1"
                /codon_start=1
                /transl_table=11
                /product="hypothetical 17.1 kDa protein in modC-bioA
                intergenic region"
                /protein_id="AAG60577.1"
                /db_xref="GI:12620125"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDN1SPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKAG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
gene            complement(587. .1876)
                /gene="bioA"
CDS             complement(587. .1876)
                /gene="bioA"
                /note="7,8-diaminopelargonic acid
                synthetase-aminotransferase"
                /codon_start=1
                /transl_table=11
                /product="DAPA-aminotransferase BioA"
                /protein_id="AAG60578.1"
                /db_xref="GI:12620126"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGKR
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LKCVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
gene            1963. .3003
                /gene="bioB"
CDS             1963. .3003
                /gene="bioB"
                /codon_start=1
                /transl_table=11
                /product="biotin synthase BioB"
                /protein_id="AAG60579.1"
                /db_xref="GI:12620127"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPNERDMPYLEQMVQGVKDLGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            3000. .4154
                /gene="bioF"
CDS             3000. .4154
                /gene="bioF"
                /note="8-amino-7-oxononanoate synthase"
                /codon_start=1
                /transl_table=11
                /product="KAPA synthetase BioF"
                /protein_id="AAG60580.1"
                /db_xref="GI:12620128"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQCAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQLVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                RASLAVIRRDEGDARREKLVSLIAHFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
gene            4141. .4896
                /gene="bioC"
CDS             4141. .4896
                /gene="bioC"
                /note="reaction step prior to synthesis of pimeloyl-CoA"
                /codon_start=1
                /transl_table=11
                /product="biotin biosynthesis protein BioC"
                /protein_id="AAG60581.1"
                /db_xref="GI:12620129"
                /translation="MATVNKQAIAAAFGRAAALYEQHADLQRQSADVLLAMLPERKYT
                HVLDAGCGPGWMTRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                TFDLAWSNLAVQWCGNLSTALRELYRVVRSKGVVAFTTLVQGSLPELHQAWQAVDERP
                HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
gene            4883. .>5526
                /gene="bioD"
CDS             4883. .>5526
                /gene="bioD"
                /note="DTB synthetase"
                /codon_start=1
                /transl_table=11
                /product="dethiobiotin synthetase BioD"
                /protein_id="AAG60582.1"
                /db_xref="GI:12620130"
                /translation="MLVSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVAS
                GSEKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESSVMS
                SGLRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAML
                TAQAIQHAGLTLAGWVANDVTPPGKRHAEYITTLTRMIPAPLLGEIPWLAENPE"
BASE COUNT      1274 a   1507 c   1567 g   1178 t
ORIGIN


  Query Match             93.6%;  Score 23.4;  DB 1;  Length 5526;
  Best Local Similarity   96.0%;  Pred. No. 0.58;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          ||||||||||||||||||||| |||
Db   2061 GCATTTCGATCCTCGTCAGGTCCAG 2085


RESULT  16
AF250768
LOCUS       AF250768     4000 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosFS1 BioA-like protein (bioA) gene, partial
            cds; BioB-like protein (bioB) and KAPA-synthetase (bioF) genes,
            complete cds; and biotin biosynthesis protein BioC (bioC) gene,
            partial cds.
ACCESSION   AF250768
VERSION     AF250768.1  GI:12620095
KEYWORDS    .
SOURCE      uncultured bacterium pCosFS1.
  ORGANISM  uncultured bacterium pCosFS1
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 4000)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 4000)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .4000
                     /organism="uncultured bacterium pCosFS1"
                     /db_xref="taxon:143795"
                     /clone="pCosFS1"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     gene            complement(<1. .1177)
                     /gene="bioA"
     CDS             complement(<1. .1177)
```

```
                /gene="bioA"
                /codon_start=1
                /transl_table=11
                /product="BioA-like protein"
                /protein_id="AAG60558.1"
                /db_xref="GI:12620098"
                /translation="MSVTDSDLEFDQRHIWHPYTSMSHPLPCYPVEAASGVELQLADG
                RRLIDGMSSWWAAIHGYNHPQLNLAATQQLEKMSHVMFGGITHPAANELCRKVVAMTP
                EALQCVFLADSGSVAVEVSLKMALQYWQARGERRQRILTLRHGYHGDTFGAMSVCDPD
                NSMHSLYQGYLAPHLFATAPQCGFDQAWDENDIAPFAALLEQHAGEVAAVILEPVVQG
                AGGMRIYHPTYLERVRELCDRYQVLLIADEIATGFGRTGKLFACEHAEVVPDILCLGK
                ALTGGYMTLSATLTTRHVAETISNGAAGCFMHGPTFMGNPLACAVANASLSLLAENRW
                QQQVAAIETQLKQQLLPLKMQPKVADVRVLGAIGVVEMREPVDMAQLQRGFVERGVWI
                "
     gene       1265. .2305
                /gene="bioB"
     CDS        1265. .2305
                /gene="bioB"
                /codon_start=1
                /transl_table=11
                /product="BioB-like protein"
                /protein_id="AAG60559.1"
                /db_xref="GI:12620099"
                /translation="MMADRIHWTVGLAQTLFDKPLLELLFEAQTVHRQHFDPRQVQVS
                TLLSIKTGACPEDCKYCPQSSRYKTGLESERLMQVEQVLESARKAKANGSTRFCMGAA
                WKNPHERDMPYLQQMVQGVKAMGMETCMTLGTLDGSQAERLAEAGLDYYNHNLDTSPE
                FYGNIITTRSYQERLDTLGKVRGAGIKVCSGGIVGLGETVXXRAGLLVQLANLPTPPE
                SVPINMLVKVKGTPLADNDDVDPFDFIRTIAVARIMMPSSYVRLSAGREQMNEQTQAM
                CFMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTETEHGDNQQQEVLAARLM
                QADTAEFYNAAL"
     gene       2305. .3456
                /gene="bioF"
     CDS        2305. .3456
                /gene="bioF"
                /note="7-keto-8-aminopelargonic acid synthetase; similar
                to BioF"
                /codon_start=1
                /transl_table=11
                /product="KAPA-synthetase"
                /protein_id="AAG60556.1"
                /db_xref="GI:12620096"
                /translation="MSWQQRIEQALEERXQSAAYRHRLVIKGGNGRYLQLNDQRYLNF
                SGNDYLGLSQNAQVIAAWQQGAQQYGVGSGGSGHVTGFCTAHQALEQQLAQWLGYPRS
                LLFISGYAANQALLAALMQGEDRILADRLSHASLLEAAAHSPAQLRRFRHNQPQALAD
                LLAKPCSGQTLAVTEGVFSMDGDSAPLAELYRLTRDAGAWLMVDDAHGIGVQGEQGRG
                SCWQQGVRPELLVVTFGKAFGLSGAAVLCDESTAEYLLQFARHLIYSTAMPPAQACAL
                QAALHCIQQGDELRTRLNTNIQRFRQGAAQLSLTLTESLTAIQPLLVGDNQRALDLAQ
                HLRGQGLWVSAIRPPTVPPGGARLRITLTAAHQPQDIDRLLEVLHDVSV"
     gene       3440. .>4000
                /gene="bioC"
     CDS        3440. .>4000
                /gene="bioC"
                /note="similar to BioC; probably reaction step prior to
                synthesis of pimeloyl-CoA"
                /codon_start=1
                /transl_table=11
                /product="biotin biosynthesis protein BioC"
                /protein_id="AAG60557.1"
                /db_xref="GI:12620097"
                /translation="MTSAFEQVNKQAVASAFSRAAVGYDAAAVLQRDVGERLLGMGQE
                HQGEQLLDAGCGTGYFSRRWRELGKQVTALDLAPGMLAFARQQQAADHYLLGDIENVP
                LGDAAVDISFSSLVVQWCSDLPRALAELYRVTRPGGVILFSTLAEGSLHELGDAWQQV
                DGERHVNDFLPLAQIEAACAGYRHHLQ"
BASE COUNT      794 a   1164 c   1236 g    802 t      4 others
ORIGIN


  Query Match             87.2%;  Score 21.8;  DB 1;  Length 4000;
  Best Local Similarity   92.0%;  Pred. No. 3.6;
  Matches   23;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||||| ||||| ||||||||||||

Db    1366 GCATTTTGATCCGCGTCAGGTGCAG 1390


RESULT  17
ECOBIOB
LOCUS       ECOBIOB       128 bp    DNA             BCT       03-JAN-1995
DEFINITION  Escherichia coli biotin (bioB) gene, early terminator region.
ACCESSION   M27731
VERSION     M27731.1  GI:341755
KEYWORDS    bioB gene; biotin.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 128)
  AUTHORS   Nath,S.K.
  TITLE     Attenuation of transcription of biotin genes in Escherichia coli
  JOURNAL   Can. J. Microbiol. 34 (12), 1288-1296 (1988)
  MEDLINE   89167942
FEATURES             Location/Qualifiers
     source          1. .128
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     gene            4. .128
                     /gene="bioB"
     -10_signal      4. .9
                     /gene="bioB"
     mRNA            20. .>128
                     /gene="bioB"
     terminator      103. .121
                     /gene="bioB"
                     /note="early terminator"
BASE COUNT       31 a     28 c     34 g     35 t
ORIGIN


  Query Match             84.0%;  Score 21;  DB 1;  Length 128;
  Best Local Similarity   100.0%;  Pred. No. 7.8;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggt 21
           |||||||||||||||||||||

Db      59 GCATTTCGATCCTCGTCAGGT 79


RESULT  18
SMABIO
LOCUS       SMABIO       7215 bp    DNA             BCT       04-FEB-1999
DEFINITION  S. marcescens biotin operon, complete sequence.
ACCESSION   D17468
VERSION     D17468.1  GI:402530
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-aminopelargonic acid synthetase; DAPA aminotransferase;
            DTB synthetase; KAPA synthetase; biotin operon; biotin synthetase;
            dethiobiotin synthetase.
SOURCE      Serratia marcescens (strain:Sr41) DNA.
  ORGANISM  Serratia marcescens
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Serratia.
REFERENCE   1  (bases 1 to 7215)
  AUTHORS   Sakurai,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1993) to the DDBJ/EMBL/GenBank databases. Naoki
            Sakurai, Tanabe Seiyaku Co., Ltd., Res Lab of Applied Biochemistry;
            2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan
            (E-mail:GFC01101@niftyserve.or.jp; nsakurai@ddbj.nig.ac.jp,
            Tel:048-433-2545, Fax:048-433-2540)
REFERENCE   2  (bases 1 to 7215)
  AUTHORS   Sakurai,N., Imai,Y., Akatsuka,H., Kawai,E., Komatsubara,S. and
            Tosa,T.
  TITLE     Complete nucleotide sequence of biotin operon of Serratia
            marcescens
```

```
  JOURNAL   Unpublished (1993)
COMMENT     Submitted (25-Aug-1993) to DDBJ by:
            Naoki Sakurai
            Res. Lab. of Applied Biochemistry
            Tanabe Seiyaku Co., Ltd.
            2-50 Kawagishi-2-chome
            Toda, Saitama 335
            Japan
            Phone: 048-433-2545
            Email: nsakurai@ddbj.nig.ac.jp
            Fax:   048-433-2540.
FEATURES             Location/Qualifiers
     source          1. .7215
                     /organism="Serratia marcescens"
                     /strain="Sr41"
                     /db_xref="taxon:615"
     gene            complement(1353. .2630)
                     /gene="bioA"
     CDS             complement(1353. .2630)
                     /gene="bioA"
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase
                     (DAPA aminotransferase)"
                     /protein_id="BAA04284.1"
                     /db_xref="GI:402531"
                     /translation="MSVTASDLAFDQRHIWHPYTSMSRPLPCYPIESASGVELQLADG
                     RSLVDGMSSWWAAIHGYNHPRLNQAASQQLEKMSHVMFGGITHPAAISLCRRLVEMTP
                     EALQCVFLADSGSVAVEVSLKMALQYWQARGERRQRILTLRHGYHGDTFGAMSVCDPD
                     NSMHSLYQGYLAPHLFATAPQCRFDEEWREEDIAPFAALLEQHAGEVAAVILEPVVQG
                     AGGMRIYHPTYLKRVRDVVRVAIKLLLIADEIATGFGRTGKLFACEHAQVVPDILCLG
                     KALTGGYMTLSATLTTRHVAETISNGAAGCFMHGPTFMGNPLACAVADASLALLAENR
                     WQAQVSAIETQLKRELLPLGRCRRWPHVRVLGAIGVVEMREPVDVAELQRGFVERGVW
                     IRPFGKLIYLMPPYIIEAEQLSRLTAAVADAAR"
     gene            2716. .3756
                     /gene="bioB"
     CDS             2716. .3756
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="BAA04285.1"
                     /db_xref="GI:402532"
                     /translation="MMADRIHWTVGQAQALFDKPLLELLFEAQTVHRQHFDPRQVQVS
                     TLLSIKTGACPEDCKYCPQSPRYKTGLESERLMQVEQVLESARKAKANGSTRFCMGAA
                     WKNPHERDMPYLQQMVQGVKAMGMETCMTLGTLDGTQAERLAEAGLDYYNHNLDTSPE
                     FYGSIITTRSYQERLDTLDKVRDAGIKVCSGGIVGLGETVRDRAGLLVQLANLPKPPE
                     SVPINMLVKVKGTPLADNDDVDPFDFIRTIAVARIMMPSSYVRLSAGREQMNEQTQAM
                     CFMAGANSIFYGGCKLLTTPNPEEDKDLQLFRKLGLNPQQTATEHGDNQQQQVLAKQLL
                     NADTAEFYNAAP"
     gene            3756. .4904
                     /gene="bioF"
     CDS             3756. .4904
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-aminopelargonic acid synthetase (KAPA
                     synthetase)"
                     /protein_id="BAA04286.1"
                     /db_xref="GI:402533"
                     /translation="MSWQQRIEQALAERRLNAAYRRRQTTEGGNGRQIRLGDRLYLNF
                     SGNDYLGLSQDARVIAAWQQGARYGVGNGGSGHVTGFSAAHQALEEQLAAWLGYPRAL
                     LFISYAANQAVLAALMQKGDRILADRLSHASLLEAAGAVAGRRAPVPAQSTAGLARIC
                     LAKPCDGQRLAVTEGLFSMDGDGAPLAELHRLTRAAGAWLMVDDAHGIGVRGEQGRGS
                     CWQQGVRPELLVATFGKAFGVSGAAVLCDEATAEYLLQFARHLIYSTAMPPAQACALQ
                     AALARIREGDDLRARLQDNIRRFRQGAAPLALTLTDSDTAIQPLLVGDNQRALDLATR
                     LRECGLWVSAIRPPTVPPGGARLRLLLTAAHQSQDIDRLLEVLNDVSQ"
     gene            4888. .5655
                     /gene="bioC"
     CDS             4888. .5655
                     /gene="bioC"
                     /function="synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="the product of bioC"
                     /protein_id="BAA04287.1"
                     /db_xref="GI:402534"
                     /translation="MTSANDTVNKQAVASAFSRAAGSYDAAAALQRDVGERLLGMGSS
                     HPGEQLLDAGCGTGYFSRMWRERGKRVTALDLAPGMLDVARQRQAAHHYLLGDIEQVP
                     LPDAAMDICFSSLVVQWCSDLPAALAELYRVTRPGGVILFSTLAAGSLQELGDAWQQV
                     DGERHVNAFLPLTQIRTACAAYRHELVTELRTLNYPDVMTLMRSLKGIGATHLHQGRE
                     GGLMSRGRLAALQAAYPCRQGQFPLSYHLAYGVIYRE"
     gene            5648. .6331
                     /gene="bioD"
     CDS             5648. .6331
                     /gene="bioD"
                     /function="conversion from 7,8-diaminopelargonic acid into
                     dethiobiotin"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase(DTB synthetase)"
                     /protein_id="BAA04288.1"
                     /db_xref="GI:912453"
                     /translation="MSKRWFVTGTDTEVGKSVASSALLQAANRAGYPSAGYKPVASGS
                     EMTAEGLRNGDALALQANSGVALDYDEVNPYVFAEPTSPHIVSADEGRPIDAARLSDG
                     LRRLEQRADWVLVEGAGGWFTPLSAEYTFADWVRQEQLPVILVVGIKLGCINHAVLTA
                     QAVQQAGLTLAGWIANDVGGAPGRRHQEYLATLRRMLPRRCWAKSRTCRRPERAPLGQ
                     YLDISLLAQ"
BASE COUNT     1347 a    2225 c    2367 g    1276 t
ORIGIN


  Query Match             80.8%;  Score 20.2;  DB 1;  Length 7215;
  Best Local Similarity   88.0%;  Pred. No. 23;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           ||| ||||| || ||||||||||||
Db    2817 GCACTTCGACCCGCGTCAGGTGCAG 2841


RESULT  19
EHBIOTOP2
LOCUS       EHBIOTOP2    1041 bp    DNA             BCT       04-NOV-1996
DEFINITION  Erwinia herbicola biotin synthetase (bioB) gene, complete cds.
ACCESSION   U38648
VERSION     U38648.1  GI:1228112
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Erwinia herbicola strain=Eho10.
  ORGANISM  Pantoea agglomerans
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Pantoea.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Wu,C.H., Chen,H.Y. and Shiuan,D.
  TITLE     Isolation and characterization of the Erwinia herbicola bio operon
            and the sequences of the bioA and bioB genes
  JOURNAL   Gene 174 (2), 251-258 (1996)
  MEDLINE   97045821
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Shiuan,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-1995) David Shiuan. Biology, Natl. Sun Yat-Sen
            University, Kaohsiung, Taiwan, ROC
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Pantoea agglomerans"
                     /strain="Eho10"
                     /db_xref="taxon:549"
     gene            1. .1041
                     /gene="bioB"
     CDS             1. .1041
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="biotin synthetase"
                     /protein_id="AAC44534.1"
                     /db_xref="GI:1228114"
                     /translation="MAHHARWTLSQVTELFDKPLLDLLFEAQTIHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLETERLMEVEQVLDSARKAKNAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLDETQAQRLASAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLDKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSFVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDVQLFRKLGLNPQQTDVMTGDNEQQQKLEEQIFN
                     ADTDQFYNAAAL"
BASE COUNT      230 a    322 c    308 g    181 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 1;  Length 1041;
  Best Local Similarity   84.0%;  Pred. No. 1.4e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          ||| ||||| ||||| ||||| |||
Db     99 GCACTTCGACCCTCGCCAGGTTCAG 123


RESULT  20
AC067950/c
LOCUS       AC067950   101721 bp    DNA             PRI       03-JUL-2001
DEFINITION  Homo sapiens clone RP11-375B23, complete sequence.
ACCESSION   AC067950
VERSION     AC067950.5  GI:14589723
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 101721)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 101721)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 101721)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 3, 2001 this sequence version replaced gi:13518257.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------------- Project Information -----------------
            Center project name: H_NH0375B23
            ------------------------------------------------------
FEATURES             Location/Qualifiers
     source          1. .101721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-375B23"
BASE COUNT    28090 a  23017 c  22013 g  28601 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 9;  Length 101721;
  Best Local Similarity   84.0%;  Pred. No. 1.7e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||||| || ||| ||||| |||||
Db  24900 GCATTTGGAACCTGGTCAGTTGCAG 24876


RESULT  21
AF165146
LOCUS       AF165146   121524 bp    DNA             HTG       07-JUN-2001
DEFINITION  Homo sapiens chromosome 8 clone CTA-397H3 map 8q12-8q13, ***
            SEQUENCING IN PROGRESS ***, 7 unordered pieces.
ACCESSION   AF165146
VERSION     AF165146.3  GI:14327840
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 121524)
  AUTHORS   Schilhabel,M.B., Baumgart,C., Blechschmidt,K., Dette,M., Jahn,N.,
            Lehmann,R., Menzel,U., Polley,A., Reichwald,K., Schudy,A.,
            Siddiqui,R., Taudien,S., Wen,G., Schlegelberger,B., Siebert,R.,
            Rosenthal,A. and Platzer,M.
  TITLE     Chromosome 8 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 121524)
  AUTHORS   Schudy,A., Blechschmidt,K., Schillhabel,M., Baumgart,C., Menzel,U.,
            Weber,J., Schattevoy,R. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT     On Jun 7, 2001 this sequence version replaced gi:8151945.
            -------------- Genome Center
                 Center: Insitute of Molecular Biotechnoloy
                 Center code: IMB
                 Web site: http://genome.imb-jena.de/
                 Contact: gscj-submit@genome.imb-jena.de
            -------------- Project Information
                 Center project name: H171
                 Center clone name: CTA-397H3
            -------------- Summary Statistics
                 Sequencing vector: M13; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990329
                 Consensus quality: 115981 bases at least Q40
                 Consensus quality: 117981 bases at least Q30
                 Consensus quality: 119332 bases at least Q20
                 Quality coverage: 9.57 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1187: contig of 1187 bp in length
            *     1188     1287: gap of unknown length
            *     1288     3530: contig of 2243 bp in length
            *     3531     3630: gap of unknown length
            *     3631    21624: contig of 17994 bp in length
            *    21625    21724: gap of unknown length
            *    21725    43656: contig of 21932 bp in length
            *    43657    43756: gap of unknown length
            *    43757    61383: contig of 17627 bp in length
            *    61384    61483: gap of unknown length
            *    61484    85318: contig of 23835 bp in length
            *    85319    85418: gap of unknown length
            *    85419   121524: contig of 36106 bp in length.
FEATURES             Location/Qualifiers
     source          1. .121524
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
```

```
                     /map="8q12-8q13"
                     /clone="CTA-397H3"
BASE COUNT    34959 a  25555 c  25579 g  34831 t    600 others
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 2;  Length 121524;
  Best Local Similarity   84.0%;  Pred. No. 1.7e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           ||||||| |||||   ||||||||
Db   74081 GCATTTCGCTCCTCAGGAGGTGCAG 74105


RESULT  22
AC021450
LOCUS       AC021450   176494 bp    DNA             HTG       04-APR-2000
DEFINITION  Homo sapiens clone RP11-65E19, WORKING DRAFT SEQUENCE, 16 unordered
            pieces.
ACCESSION   AC021450
VERSION     AC021450.3  GI:7407964
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176494)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-65E19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176494)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 4, 2000 this sequence version replaced gi:6814993.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5159
                Center clone name: 65_E_19
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 167673 bases at least Q40
                Consensus quality: 171804 bases at least Q30
                Consensus quality: 173582 bases at least Q20
                Insert size: 180000; agarose-fp
                Insert size: 174994; sum-of-contigs
                Quality coverage: 4.7 in Q20 bases; agarose-fp
                Quality coverage: 4.8 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1      949: contig of 949 bp in length
            *       950 1049: gap of      100 bp
            *      1050     2544: contig of 1495 bp in length
            *      2545 2644: gap of      100 bp
            *      2645     5857: contig of 3213 bp in length
            *      5858 5957: gap of      100 bp
            *      5958     9641: contig of 3684 bp in length
            *      9642 9741: gap of      100 bp
            *      9742    14928: contig of 5187 bp in length
            *     14929 15028: gap of      100 bp
            *     15029    20347: contig of 5319 bp in length
            *     20348 20447: gap of      100 bp
            *     20448    27447: contig of 7000 bp in length
            *     27448 27547: gap of      100 bp
            *     27548    39249: contig of 11702 bp in length
            *     39250 39349: gap of      100 bp
            *     39350    46799: contig of 7450 bp in length
            *     46800 46899: gap of      100 bp
            *     46900    56544: contig of 9645 bp in length
            *     56545 56644: gap of      100 bp
            *     56645    65562: contig of 8918 bp in length
            *     65563 65662: gap of      100 bp
            *     65663    81928: contig of 16266 bp in length
            *     81929 82028: gap of      100 bp
            *     82029    97874: contig of 15846 bp in length
            *     97875 97974: gap of      100 bp
            *     97975   118702: contig of 20728 bp in length
            *    118703 118802: gap of      100 bp
            *    118803   142786: contig of 23984 bp in length
            *    142787 142886: gap of      100 bp
            *    142887   176494: contig of 33608 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176494
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-65E19"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .949
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    1050. .2544
                     /note="assembly_fragment"
     misc_feature    2645. .5857
                     /note="assembly_fragment"
     misc_feature    5958. .9641
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    9742. .14928
                     /note="assembly_fragment"
     misc_feature    15029. .20347
                     /note="assembly_fragment"
     misc_feature    20448. .27447
                     /note="assembly_fragment"
     misc_feature    27548. .39249
                     /note="assembly_fragment"
     misc_feature    39350. .46799
                     /note="assembly_fragment"
     misc_feature    46900. .56544
                     /note="assembly_fragment"
     misc_feature    56645. .65562
```

```
                    /note="assembly_fragment"
     misc_feature    65663. .81928
                    /note="assembly_fragment"
     misc_feature    82029. .97874
                    /note="assembly_fragment"
     misc_feature    97975. .118702
                    /note="assembly_fragment"
     misc_feature    118803. .142786
                    /note="assembly_fragment"
     misc_feature    142887. .176494
                    /note="assembly_fragment"
BASE COUNT    55484 a  31949 c  31595 g  55963 t   1503 others
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 2;  Length 176494;
  Best Local Similarity   84.0%;  Pred. No. 1.7e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          ||| ||| ||||| | |||||||||
Db  89764 GCACTTCAATCCTGGCCAGGTGCAG 89788


RESULT  23
AC017911
LOCUS       AC017911   118100 bp    DNA             HTG       09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017911
VERSION     AC017911.1  GI:6553279
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 118100)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212606 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .118100
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    32824 a  26059 c  26550 g  32667 t
ORIGIN


  Query Match             73.6%;  Score 18.4;  DB 2;  Length 118100;
  Best Local Similarity   95.0%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcagg 20
          ||||||||||||||||| |||
Db  63425 GCATTTCGATCCTCGTAAGG 63444


RESULT  24
AC007476/c
LOCUS       AC007476   171705 bp    DNA             INV       17-MAR-2001
DEFINITION  Drosophila melanogaster, chromosome 2R, region 49A-49B, BAC clone
            BACR37P06, complete sequence.
ACCESSION   AC007476
VERSION     AC007476.5  GI:13374640
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 171705)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2R, region 49A-49B
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171705)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Mar 17, 2001 this sequence version replaced gi:5670614.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .171705
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="49A-49B"
                     /clone="BACR37P06 (D594)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    47654 a  38437 c  38157 g  47457 t
ORIGIN


  Query Match             73.6%;  Score 18.4;  DB 3;  Length 171705;
  Best Local Similarity   95.0%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcagg 20
          ||||||||||||||||| |||
Db 106124 GCATTTCGATCCTCGTAAGG 106105


RESULT  25
AE003822/c
LOCUS       AE003822   266133 bp    DNA             INV       04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386047 section 21
```

```
          of 52, complete sequence.
ACCESSION AE003822 AE002787
VERSION   AE003822.2  GI:10727606
KEYWORDS  HTG.
SOURCE    fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1  (bases 1 to 266133)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE 2  (bases 1 to 266133)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT   On Oct 9, 2000 this sequence version replaced gi:7303437.
FEATURES             Location/Qualifiers
     source          1. .266133
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
     mRNA            join(392. .425,1007. .1155,5710. .6105)
                     /gene="Cam"
                     /product="CT41657"
                     /db_xref="FLYBASE:FBan0008472"
                     /db_xref="FLYBASE:FBgn0000253"
     gene            <392. .>15192
                     /gene="Cam"
                     /note="CG8472"
                     /map="48F5-48F5"
                     /db_xref="FLYBASE:FBan0008472"
                     /db_xref="FLYBASE:FBgn0000253"
     mRNA            join(405. .425,1007. .1155,5710. .5884,10160. .10402,
                     14732. .15192)
                     /gene="Cam"
                     /product="CT41655"
                     /db_xref="FLYBASE:FBan0008472"
                     /db_xref="FLYBASE:FBgn0000253"
     mRNA            join(596. .619,1004. .1155,5710. .5884,10160. .10402,
                     14732. .14876)
                     /gene="Cam"
                     /product="CT41659"
                     /db_xref="FLYBASE:FBan0008472"
                     /db_xref="FLYBASE:FBgn0000253"
     CDS             join(1153. .1155,5710. .5922)
                     /gene="Cam"
                     /note="Cam gene product [alt 1]"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0008472"
                     /db_xref="FLYBASE:FBgn0000253"
                     /protein_id="AAF58544.1"
                     /db_xref="GI:7303488"
                     /translation="MADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNP
                     TEAELQDMINEVDADGKLWRIVGGNTL"
     CDS             join(1153. .1155,5710. .5884,10160. .10402,14732. .14760)
                     /gene="Cam"
                     /note="Cam gene product [alt 4]"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0008472"
                     /db_xref="FLYBASE:FBgn0000253"
                     /protein_id="AAF58542.1"
                     /db_xref="GI:7303486"
                     /translation="MADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNP
                     TEAELQDMINEVDADGNGTIDFPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGFIS
                     AAELRHVMTNLGEKLTDEEVDEMIREADIDGDGQVNYEEFVTMMTSK"
     CDS             join(1153. .1155,5710. .5884,10160. .10402,14732. .14760)
                     /gene="Cam"
                     /note="Cam gene product [alt 2]"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0008472"
                     /db_xref="FLYBASE:FBgn0000253"
                     /protein_id="AAF58541.1"
                     /db_xref="GI:7303485"
                     /translation="MADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNP
                     TEAELQDMINEVDADGNGTIDFPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGFIS
                     AAELRHVMTNLGEKLTDEEVDEMIREADIDGDGQVNYEEFVTMMTSK"
     mRNA            join(5527. .5642,5710. .5884,10160. .10402,14732. .15192)
                     /gene="Cam"
                     /product="CT24787"
                     /db_xref="FLYBASE:FBan0008472"
                     /db_xref="FLYBASE:FBgn0000253"
     CDS             join(5640. .5642,5710. .5884,10160. .10402,14732. .14760)
                     /gene="Cam"
                     /note="Cam gene product [alt 3]"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0008472"
                     /db_xref="FLYBASE:FBgn0000253"
                     /protein_id="AAF58543.1"
                     /db_xref="GI:7303487"
                     /translation="MADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNP
                     TEAELQDMINEVDADGNGTIDFPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGFIS
                     AAELRHVMTNLGEKLTDEEVDEMIREADIDGDGQVNYEEFVTMMTSK"
     mRNA            join(<21254. .21464,21584. .>21600)
                     /gene="CG13166"
                     /product="CT32407"
                     /db_xref="FLYBASE:FBan0013166"
                     /db_xref="FLYBASE:FBgn0040758"
                     /evidence=not_experimental
     gene            <21254. .>21600
                     /gene="CG13166"
                     /map="48F5-48F6"
                     /db_xref="FLYBASE:FBan0013166"
                     /db_xref="FLYBASE:FBgn0040758"
                     /evidence=not_experimental
     CDS             join(21254. .21464,21584. .21600)
```

```
                     /gene="CG13166"
                     /note="CG13166 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0013166"
                     /db_xref="FLYBASE:FBgn0040758"
                     /evidence=not_experimental
                     /protein_id="AAF58540.1"
                     /db_xref="GI:7303484"
                     /translation="MSERGSGGAVTGGRHARVSPTPGRGSKIAYTPCPRSEPPQDLAL
                     AVPPPPRNCLAIPMSSLQTANGPGTKELHNP"
     mRNA            join(<34383. .34566,36219. .36635,37304. .37431,37529. .37772,
                     37837. .37951,38022. .38163,38232. .38305,38481. .>38679)
                     /gene="CG13165"
                     /product="CT32406"
                     /db_xref="FLYBASE:FBan0013165"
                     /db_xref="FLYBASE:FBgn0033707"
                     /evidence=not_experimental
     gene            <34383. .>38679
                     /gene="CG13165"
                     /map="48F5-48F8"
                     /db_xref="FLYBASE:FBan0013165"
                     /db_xref="FLYBASE:FBgn0033707"
                     /evidence=not_experimental
     CDS             join(34383. .34566,36219. .36635,37304. .37431,37529. .37772,
                     37837. .37951,38022. .38163,38232. .38305,38481. .38679)
                     /gene="CG13165"
                     /note="CG13165 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0013165"
                     /db_xref="FLYBASE:FBgn0033707"
                     /evidence=not_experimental
                     /protein_id="AAF58539.1"
                     /db_xref="GI:7303483"
                     /translation="MFRKNFREGASKSGGGGGKLQSKANRTKRSRDMGLVQQPEEIHY
                     RTHLFFSPNRPGYDVGEERCSALSGIPSTPLSSTAPSLALPGNLPYELDMPQPLVDRR
                     PSVSLMRWSSSGPGTECNSNSNGNSASNSLIRLQQQTQQQMQQQLQQQLQQQLQQQMQ
                     QQQQQHLLATATATTLGYADATVATATTTLLAAAAAATRQVNFGGGANSGVTRAEPIS
                     LATLDRDCFIIPVHSVDRFLPAGIPLPALSADGKATSPLSVLEVSDPKLCILVHLMSP
                     LEAIDPVMESPLAHPLLKQRSIAAELVTEVQQANTAVGGMLLANMEKSSEFPFISYYL
                     1NTLQTDPSSFYAGLRVSSLSKFEPKALKYTAAHTLDLYSEVAAICRPPLVLPSNEVG
                     SFKKSQTAATGYIISVFKVFEGDDGERFEKNWLYWTGARMLYRYLPRAAGLRRIALHK
                     STSQKGDKMYLLVCECAELLKDISLAAFLIPALRARLCGYTGLYRPIQAF"
     mRNA            complement(join(39898. .40041,40104. .40842,40899. .41276,
                     41341. .41560,41620. .41902,41964. .42200,43133. .>43167))
                     /gene="CG8850"
                     /product="CT25422"
                     /db_xref="FLYBASE:FBan0008850"
                     /db_xref="FLYBASE:FBgn0033708"
     gene            complement(<39898. .>43167)
                     /gene="CG8850"
                     /map="48F5-48F8"
                     /db_xref="FLYBASE:FBan0008850"

  Query Match             73.6%;  Score 18.4;  DB 3;  Length 266133;
  Best Local Similarity   95.0%;  Pred. No. 2.2e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcagg 20
           |||||||||||||||| |||
Db 158787 GCATTTCGATCCTCGTAAGG 158768


RESULT  26
AC023391/c
LOCUS       AC023391   154270 bp    DNA             HTG       25-JUN-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-147K21 map 4, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC023391
VERSION     AC023391.4  GI:8705090
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154270)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-147K21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154270)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 25, 2000 this sequence version replaced gi:7677804.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5405
                Center clone name: 147_K_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 149061 bases at least Q40
                Consensus quality: 151960 bases at least Q30
                Consensus quality: 152881 bases at least Q20
                Insert size: 154000; agarose-fp
                Insert size: 153470; sum-of-contigs
                Quality coverage: 4.7 in Q20 bases; agarose-fp
                Quality coverage: 4.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1004: contig of 1004 bp in length
            *    1005 1104: gap of      100 bp
            *    1105     7217: contig of 6113 bp in length
            *    7218 7317: gap of      100 bp
            *    7318    18994: contig of 11677 bp in length
            *   18995 19094: gap of      100 bp
            *   19095    35888: contig of 16794 bp in length
            *   35889 35988: gap of      100 bp
            *   35989    52052: contig of 16064 bp in length
            *   52053 52152: gap of      100 bp
            *   52153    69573: contig of 17421 bp in length
```

```
            *    69574 69673: gap of      100 bp
            *    69674    93465: contig of 23792 bp in length
            *    93466 93565: gap of      100 bp
            *    93566   116170: contig of 22605 bp in length
            *   116171 116270: gap of      100 bp
            *   116271   154270: contig of 38000 bp in length.
FEATURES             Location/Qualifiers
     source          1. .154270
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-147K21"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1004
                     /note="assembly_fragment"
     misc_feature    1105. .7217
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    7318. .18994
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    19095. .35888
                     /note="assembly_fragment"
     misc_feature    35989. .52052
                     /note="assembly_fragment"
     misc_feature    52153. .69573
                     /note="assembly_fragment"
     misc_feature    69674. .93465
                     /note="assembly_fragment"
     misc_feature    93566. .116170
                     /note="assembly_fragment"
     misc_feature    116271. .154270
                     /note="assembly_fragment"
BASE COUNT    52075 a  27858 c  27285 g  46252 t    800 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 2;  Length 154270;
  Best Local Similarity   87.0%;  Pred. No. 2.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgca 24
          ||||| | |||||||||| |||||
Db  39854 CATTTGGCTCCTCGTCATGTGCA 39832


RESULT  27
AB061540
LOCUS       AB061540     1946 bp    DNA             BCT       16-MAY-2001
DEFINITION  Desulfovibrio longus dsrA, dsrB genes for sulfite reductase alpha
            and beta subunits, partial cds.
ACCESSION   AB061540
VERSION     AB061540.1  GI:14090316
KEYWORDS    .
SOURCE      Desulfovibrio longus (strain:DSM6739) DNA.
  ORGANISM  Desulfovibrio longus
            Bacteria; Proteobacteria; delta subdivision; Desulfovibrio.
REFERENCE   1  (sites)
  AUTHORS   Kondo,R.
  TITLE     Nucleotide sequences of dissimilatory sulfite reductase genes from
            sulfate-reducing bacteria
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1946)
  AUTHORS   Kondo,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2001) Ryuji Kondo, Fukui Prefectural University,
            Department of Marine Bioscience; 1-1 Gakuen, Obama, Fukui 917-0003,
            Japan (E-mail:rykondo@fpu.ac.jp, Tel:81-770-52-6300(ex.1705),
            Fax:81-770-6003)
FEATURES             Location/Qualifiers
     source          1. .1946
                     /organism="Desulfovibrio longus"
                     /strain="DSM6739"
                     /db_xref="taxon:889"
     gene            1. .1134
                     /gene="dsrA"
     CDS             <1. .1134
                     /gene="dsrA"
                     /codon_start=1
                     /transl_table=11
                     /product="sulfite reductase alpha subunit"
                     /protein_id="BAB55573.1"
                     /db_xref="GI:14090317"
                     /translation="THWKHGGIVGVFGYGGGVIGRYCDQPEMFPGVAHFHTVRVNQPS
                     GKFYTTEFLRQLCDLWEFRGSGLTNMHGATGDIVFIGTTTPQLEEIFFELTHNLNTDL
                     GGSGSNLRTPATCMGMSRCEFACYDTQALCYDMTNEYQDELHRPAFPYKFKFKFDGCP
                     NGCVASIARSDFSVIGTWRDAIRIDQKAVAGYVGGEIAPNAGAHAGRDWGKFDIQKEV
                     IDLCPSECMSYEGGKLAINDKECVRCMHCINTMPQALRIGNDRGASIFVGAKAPILDG
                     AQMGSLLVPFVKVDADNEYREIKDVVEGIWDLWMEEGKNRERLGETMKRMGFNKIIEA
                     AGVPFDSRQVQEPRTNPYIFWKAEDVEGGWDRDISDYRKRHKL"
     gene            1150. .1946
                     /gene="dsrB"
     CDS             1150. .>1946
                     /gene="dsrB"
                     /codon_start=1
                     /transl_table=11
                     /product="sulfite reductase beta subunit"
                     /protein_id="BAB55574.1"
                     /db_xref="GI:14090318"
                     /translation="MAFISSGYNPAKPMENRITDIGPRHFTDFLPPVIKKNFGKWLYH
                     EILEPGVLMHKAESGDEVYTVRCGGARLMTVTHIREICDVADKFADGHLRFTTRNNIE
                     FMLESKDKALELKKYLNAQKFEGGSYKFPVGGTGAGVTNIVHTQGWVHCHTPATDASG
                     TVKATMDVVFEEFQNMRLPAPVRISMACCLNMCGAVHCSDIAILGIHRKPPMIDNKIL
                     DNLCEVPLAVAACPTGAIRPTKIEIDGQKYNTVAVKEERCMFCGNCYT"
BASE COUNT      403 a    661 c    541 g    341 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 1;  Length 1946;
  Best Local Similarity   90.5%;  Pred. No. 3.5e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      5 ttcgatcctcgtcaggtgcag 25
          |||||| | ||||||||||||
Db   1015 TTCGATTCGCGTCAGGTGCAG 1035


RESULT  28
AC091701/c
LOCUS       AC091701   144483 bp    DNA             HTG       17-JUL-2001
DEFINITION  Trypanosoma brucei chromosome VIII clone RPCI93-26N11, ***
            SEQUENCING IN PROGRESS ***, 1 ordered pieces.
ACCESSION   AC091701
VERSION     AC091701.2  GI:14285330
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 144483)
  AUTHORS   El-Sayed,N.M., Ghedin,E., Song,J., Larkin,C., Wanless,D., Jones,K.,
            Peterson,J., Hou,L., Zhao,H., Mason,T., Militscher,J., Pai,G., Van
            Aken,S., Utterback,T., Khalak,H.G., Gerard,C., Leech,V., Ullu,E.,
            Melville,S., White,O., Adams,M.D., Donelson,J.E. and Fraser,C.M.
  TITLE     Trypanosoma brucei GUTat10.1 RPCI93-26N11 BAC genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 144483)
  AUTHORS   El-Sayed,N.M., Khalak,H. and Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On Jun 4, 2001 this sequence version replaced gi:14161709.
            * NOTE: This is a 'working draft' sequence. It currently
```

```
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   144483: contig of 144483 bp in length.
FEATURES             Location/Qualifiers
     source          1. .144483
                     /organism="Trypanosoma brucei"
                     /isolate="GUTat10.1"
                     /db_xref="taxon:5691"
                     /chromosome="VIII"
                     /clone="RPCI93-26N11"
BASE COUNT    36739 a  34697 c  35464 g  37583 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 2;  Length 144483;
  Best Local Similarity   90.5%;  Pred. No. 4.2e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggt 21
          ||||||  |||||||  ||||||
Db  66403 GCATTTAGATCCTCTTCAGGT 66383


RESULT  29
AX157421/c
LOCUS       AX157421       51 bp    DNA             PAT       22-JUN-2001
DEFINITION  Sequence 749 from Patent WO0140521.
ACCESSION   AX157421
VERSION     AX157421.1  GI:14538752
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Shimkets,R.A. and Leach,M.
  TITLE     Nucleic acids containing single nucleotide polymorphisms and
            methods of use thereof
  JOURNAL   Patent: WO 0140521-A 749 07-JUN-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     misc_feature    26
                     /note="1 of 2 allelic variants (750 is other entry)
                     Accession number cg21425684"
BASE COUNT        9 a     18 c     15 g      9 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 6;  Length 51;
  Best Local Similarity   83.3%;  Pred. No. 3.8e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgca 24
          |||  ||||  ||  ||||||||||||
Db     24 GCACGTCGAACCGCGTCAGGTGCA 1


RESULT  30
AX157422/c
LOCUS       AX157422       51 bp    DNA             PAT       22-JUN-2001
DEFINITION  Sequence 750 from Patent WO0140521.
ACCESSION   AX157422
VERSION     AX157422.1  GI:14538753
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Shimkets,R.A. and Leach,M.
  TITLE     Nucleic acids containing single nucleotide polymorphisms and
            methods of use thereof
  JOURNAL   Patent: WO 0140521-A 750 07-JUN-2001;
            Curagen Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     misc_feature    26
                     /note="2 of 2 allelic variants (749 is other entry)
                     Accession number cg21425684"
BASE COUNT        9 a     18 c     14 g     10 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 6;  Length 51;
  Best Local Similarity   83.3%;  Pred. No. 3.8e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgca 24
          |||  ||||  ||  ||||||||||||
Db     24 GCACGTCGAACCGCGTCAGGTGCA 1


RESULT  31
AX185944
LOCUS       AX185944       295 bp    DNA             PAT       06-AUG-2001
DEFINITION  Sequence 1639 from Patent WO0142467.
ACCESSION   AX185944
VERSION     AX185944.1  GI:15137361
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 295)
  AUTHORS   Schlegel,R., Deeds,J., Berger,A. and Zhao,X.
  TITLE     Genes, compositions, kits, and methods for identification,
            assessment, prevention, and therapy of cervical cancer
  JOURNAL   Patent: WO 0142467-A 1639 14-JUN-2001;
            Millennium Predictive Medicine, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .295
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       82 a     62 c     58 g     93 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 6;  Length 295;
  Best Local Similarity   83.3%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgcag 25
          |||||  ||  ||||  ||||||  |||
Db    132 CATTTGGAGCCTCATCAGGTTCAG 155


RESULT  32
AX185204
LOCUS       AX185204       321 bp    DNA             PAT       06-AUG-2001
DEFINITION  Sequence 899 from Patent WO0142467.
ACCESSION   AX185204
VERSION     AX185204.1  GI:15136588
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 321)
  AUTHORS   Schlegel,R., Deeds,J., Berger,A. and Zhao,X.
  TITLE     Genes, compositions, kits, and methods for identification,
            assessment, prevention, and therapy of cervical cancer
  JOURNAL   Patent: WO 0142467-A 899 14-JUN-2001;
            Millennium Predictive Medicine, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .321
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       83 a     73 c     68 g     97 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 6;  Length 321;
  Best Local Similarity   83.3%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgcag 25
          ||||| || |||| |||||| |||
Db    159 CATTTGGAGCCTCATCAGGTTCAG 182


RESULT  33
AK024633
LOCUS       AK024633     1315 bp    mRNA            PRI       29-SEP-2000
DEFINITION  Homo sapiens cDNA: FLJ20980 fis, clone ADSU01986.
ACCESSION   AK024633
VERSION     AK024633.1  GI:10436953
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens adipose tissue cDNA to mRNA, clone_lib:ad
            clone:ADSU01986.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
            Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Tanigami,A.,
            Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hirao,M.,
            Ohmori,Y., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T.,
            Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 1315)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
            Sugano, Institute of Medical Science, University of Tokyo,
            Laboratory of Genome Structure Analysis, Human Genome Center;
            Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Departent of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1. .1315
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="ADSU01986"
                     /clone_lib="ad"
                     /tissue_type="adipose tissue"
                     /note="cloning vector pME18SFL3"
     CDS             283. .1257
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB14939.1"
                     /db_xref="GI:10436954"
                     /translation="MESLFPYGPTMPQTLAGLSPGALLQQYQQYQQNLQESLQKQQKQ
                     QQEQQQKPVQAKTSKVESDQPQNSNDASETKEDKSTATESTKEEPQLESKSADFSDTY
                     VVPFVKYEFICRKCQMMFTDEDAAVNHQKSFCYFGQPLIDPQETVLRVPVSKYQCLAC
                     DVAISGNEALSQHLQSSLHKEKTIKQAMRNAKEHVRLLPHSVCSPNPNTTSTSQSAAS
                     SNNTYPHLSCFSMKSWPNILFQASARRAASPPSSPPSLSLPSTVTSSLCSTSGVQTSL
                     PTESCSDESDSELSQKLEDLDNSLEVKAKPASGLDGNFNSIRMDMFSV"
BASE COUNT      409 a    326 c    266 g    314 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 9;  Length 1315;
  Best Local Similarity   83.3%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgcag 25
          ||| |||||||| |||| ||||
Db    115 CATGTCGATCCTATTCAGTTGCAG 138


RESULT  34
AF298117/c
LOCUS       AF298117    10728 bp    DNA             PRI       26-SEP-2000
DEFINITION  Homo sapiens homeobox protein OTX2 gene, complete cds.
ACCESSION   AF298117
VERSION     AF298117.1  GI:10304407
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 10728)
  AUTHORS   Perrault,I., Rozet,J.-M., Gerber,S., Munnich,A. and Kaplan,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2000) Inserm U393, Hopital Necker Enfants
            Malades, 149 rue De Sevres, Paris 75743 CEDEX 15, France
FEATURES             Location/Qualifiers
     source          1. .10728
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q21-q22"
     mRNA            join(<1427. .1523,2520. .2695,4528. .>5148)
                     /product="homeobox protein OTX2"
     CDS             join(1427. .1523,2520. .2695,4528. .5148)
                     /note="similar to Homo sapiens homeobox protein OTX2 in
                     Accession Number AB037505"
                     /codon_start=1
                     /product="homeobox protein OTX2"
                     /protein_id="AAG16243.1"
                     /db_xref="GI:10304408"
                     /translation="MMSYLKQPPYAVNGLSLTTSGMDLLHPSVGYPGPWASCPAATPR
                     KQRRERTTFTRAQLDVLEALFAKTRYPDIFMREEVALKINLPESRVQVWFKNRRAKCR
                     QQQQQQQNGGQNKVRPAKKKTSPAREVSSESGTSGQFTPPSSTSVPTIASSSAPVSIW
                     SPASISPLSDPLSTSSSCMQRSYPMTYTQASGYSQGYAGSTSYFGGMDCGSYLTPMHH
                     QLPGPGATLSPMGTNAVTSHLNQSPASLSTQGYGASSLGFNSTTDCLDYKDQTASWKL
                     NFNADCLDYKDQTSSWKFQVL"
BASE COUNT     2800 a   2486 c   2588 g   2854 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 9;  Length 10728;
  Best Local Similarity   83.3%;  Pred. No. 4.8e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgcag 25
          ||||| | |||||| ||| |||||
Db    306 CATTTGGGTCCTCGGCAGCTGCAG 283


RESULT  35
F6E21/c
```

```
LOCUS       F6E21       38340 bp    DNA             HTG       06-AUG-1999
DEFINITION  Arabidopsis thaliana clone F6E21, *** SEQUENCING IN PROGRESS ***,
            in unordered pieces.
ACCESSION   AL049914
VERSION     AL049914.1  GI:4884020
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 38340)
  AUTHORS   Lennard,N., Quail,M., Harris,B., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-1999) EU Arabidopsis sequencing project, The
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            For more information about this sequence or the ESSAII project, see
            MIPS http://websvr.mips.biochem.mpg.de/proj/thal/. Order of
            segments is not known; 800 n's separate segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .38340
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /clone="F6E21"
BASE COUNT    11527 a   7337 c   7836 g  11640 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 2;  Length 38340;
  Best Local Similarity   83.3%;  Pred. No. 5e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgca 24
           | |||| |||| |||||| |||||
Db   28534 GAATTTGGATCTTCGTCATGTGCA 28511


RESULT  36
AC016941
LOCUS       AC016941    72968 bp    DNA             PRI       24-JUN-2000
DEFINITION  Homo sapiens Xq28 BAC RP5-869M20 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC016941
VERSION     AC016941.9  GI:8567654
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 72968)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Brundage,E., Buhay,C.,
            Bunac,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z.,
            Cox,C., David,R., Delgado,O., Deshazo,D., Di,W., Ding,Y.,
            Domah-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C.,
            Ferraguto,D., Forcum-Tansey,J., Frantz,P., Ganesh,R., Garcia,C.,
            Gorrell,J.H., Gorrell,L.L., Guevara,W., Harris,K., Hernandez,J.,
            Hodgson,A., Hoques,M., Holloway,C., Hosak,H., Jackson,L.E.,
            Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondejewski,N., Kong,Y.,
            Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O.,
            Lozado,R.J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P.,
            Mei,G., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R.,
            Nguyen,N., Nguyen,S., Oswal,G., Pampell,L.R., Parish,B., Paxton,S.,
            Payton,B., Perez,L., Pu,L.L., Quiles,M., Rashid,N.D., Reiter,D.,
            Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H.,
            Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T.,
            Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S.,
            Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
            Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 72968)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 72968)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 72968)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 72968)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 15, 2000 this sequence version replaced gi:7114778.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
            html.

            QUALSTAT-REPORT----------
```

```
---------- Summary Statistics ----------
Contig length:                                  74593
Phrap values in estimate:                       74530
Average error rate (BCM-Phrap estimate):        0.000293652
Fraction of Phrap values less than 40 :         0.0348182
Number of consensus changing edits:             16
Number of N's in consensus :                    0


---------- Consensus changing edits ----------
Position      Original+Context          Edited+Context
5242          ctggacgtta(n)aattaaagca   ctggacgtta(a)aattaaagca
11447         caaattcttg(n)ttttttttct   caaattcttg(t)ttttttttct
33758         aaactcaatc(n)nnggattcta   aaactcaatc(c)aaggattcta
33759         aactcaatcn(n)nggattctaa   aactcaatcc(a)aggattctaa
33760         actcaatcnn(n)ggattctaag   actcaatcca(a)ggattctaag
41720         aatggaaagc(n)nctttgaaat   aatggaaagc(a)actttgaaat
41721         atggaaagcn(n)ctttgaaatg   atggaaagca(a)ctttgaaatg
43730         gaaatagaca(n)caataacatg   gaaatagaca(a)caataacatg
66538         ggtgggtgga(n)nacttgaggt   ggtgggtgga(t)cacttgaggt
66539         gtgggtggan(n)acttgaggtc   gtgggtggat(c)acttgaggtc
66607         taaaaataca(n)nnnttagcca   taaaaataca(a)aaattagcca
66608         aaaaatacan(n)nnttagccag   aaaaatacaa(a)aattagccag
66609         aaaatacann(n)nttagccagg   aaaatacaaa(a)attagccagg
66610         aaatacannn(n)ttagccaggc   aaatacaaaa(a)ttagccaggc
66653         ccagctattc(n)anaggctgag   ccagctattc(g)agaggctgag
66655         agctattcna(n)aggctgaggc   agctattcga(g)aggctgaggc


--------- Distribution of Quality < 40 Bases ----------

      1000|
       900|                                   *
       800|                                   *    *
  #    700|                                   *    *
  bases 600|                                  *    *
       500|                                   *    *
       400|                                   *    *
       300|                                   *    *
       200|                              *    *    *
       100|                   *    *    *    *    *
         0|        *    *    *    *    *    *    *
          ----------------------------------------
               5    10   15   20   25   30   35   40
                        Phrap Value Range


---------------------------------------------------------------------------
Version: 1.01  qxfo.
FEATURES             Location/Qualifiers
     source          1. .72968
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="Xq28"
                     /clone="RP5-869M20"
     repeat_region   1. .52
                     /rpt_family="GA-rich"
     misc_difference 53
                     /note="IS2 element inserted in genomic clone.  Sequence
                     removed from genomic sequence."
                     /db_xref="GI:41525"
     repeat_region   55. .810
                     /rpt_family="HERV17"
     repeat_region   738. .1488
                     /rpt_family="Harlequin"
     repeat_region   complement(929. .1011)
                     /rpt_family="LTR8"
     repeat_region   1489. .1996
                     /rpt_family="LTR10C"
     repeat_region   1997. .2165
                     /rpt_family="HERV1"
     STS             2205. .2334
                     /standard_name="DXS7085"
                     /db_xref="GDB:452929"
     misc_feature    2499. .2954
                     /note="Region similar to Hs#S886894 zi27a11.s1 Homo
                     sapiens cDNA: AA678246"
     repeat_region   3114. .3177
                     /rpt_family="GA-rich"
     repeat_region   3240. .3957
                     /rpt_family="Harlequin"
     misc_feature    3972. .4145
                     /note="Region similar to Hs#S795243 nh31d05.s1 Homo
                     sapiens cDNA: AA524937"
     repeat_region   4146. .5813
                     /rpt_family="BaEV-int"
     repeat_region   5849. .6062
                     /rpt_family="BaEV-int"
     repeat_region   6066. .6582
                     /rpt_family="LTR10C"
     repeat_region   7662. .9000
                     /rpt_family="HERVI"
     repeat_region   9006. .9470
                     /rpt_family="LTR10A"
     repeat_region   complement(9486. .9918)
                     /rpt_family="MER4-internal"
     repeat_region   9940. .10593
                     /rpt_family="L1MCb"
     repeat_region   complement(10594. .10903)
                     /rpt_family="AluY"
     repeat_region   10904. .11153
                     /rpt_family="L1MCb"
     repeat_region   complement(11142. .12690)

  Query Match             70.4%;  Score 17.6;  DB 9;  Length 72968;
  Best Local Similarity   83.3%;  Pred. No. 5.2e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgca 24
           ||||| | || |||||||||| |||
Db   51662 GCATTCCCATTCTCGTCAGGAGCA 51685


RESULT  37
AC005646
LOCUS       AC005646    86677 bp    DNA             HTG       10-DEC-1999
DEFINITION  Drosophila melanogaster chromosome 2 clone DS00968 (D404) map
            51B1-51B4 strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 3
            unordered pieces.
ACCESSION   AC005646
VERSION     AC005646.6  GI:6554236
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 86677)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 86677)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
```

```
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Dec 10, 1999 this sequence version replaced gi:5656719.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases. P1 library location:
            8-11.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      296: contig of 296 bp in length
            *      297      376: gap of unknown length
            *      377    85947: contig of 85571 bp in length
            *    85948    86027: gap of unknown length
            *    86028    86677: contig of 650 bp in length.
FEATURES             Location/Qualifiers
     source          1. .86677
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2"
                     /map="51B1-51B4"
                     /clone="DS00968 (D404)"
                     /clone_lib="P1 library, partial Sau3A in pNS582tet14Ad10"
BASE COUNT    24015 a  18183 c  18237 g  26082 t    160 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 2;  Length 86677;
  Best Local Similarity   83.3%;  Pred. No. 5.2e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgcag 25
          || || |||||| |||||| ||||
Db  60896 CAATTGGATCCTGGTCAGGAGCAG 60919


RESULT  38
LMFLCHR15_6
WPCOMMENT
  Sequence split into 8 fragments  LOCUS LMFLCHR15 Accession AL160371
     Fragment Name      Begin     End
     LMFLCHR15_0            1    110000
     LMFLCHR15_1       100001    210000
     LMFLCHR15_2       200001    310000
     LMFLCHR15_3       300001    410000
     LMFLCHR15_4       400001    510000
     LMFLCHR15_5       500001    610000
     LMFLCHR15_6       600001    710000
     LMFLCHR15_7       700001    720704
Continuation (7 of 8) of LMFLCHR15 from base 600001 (AL160371 Leishmania major chromosom


  Query Match             70.4%;  Score 17.6;  DB 2;  Length 110000;
  Best Local Similarity   83.3%;  Pred. No. 5.3e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgca 24
          ||||||| ||||||||| |  |||
Db  83492 GCATTTCCATCCTCGTCGGCGGCA 83515
```

```
RESULT  39
HSDJ335E1
LOCUS       HSDJ335E1  143805 bp    DNA             PRI       25-MAY-2000
DEFINITION  Human DNA sequence from clone RP3-335E1 on chromosome 6q16.2-21
            Contains STSs, GSSs and a CpG Island, complete sequence.
ACCESSION   AL109920
VERSION     AL109920.15  GI:7406500
KEYWORDS    HTG; CpG Island.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143805)
  AUTHORS   Frankland,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 3, 2000 this sequence version replaced gi:7378738.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP3-335E1 is from the library RPCI-3 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP3-335E1.
FEATURES             Location/Qualifiers
     source          1. .143805
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="q16.2-21"
                     /clone="RP3-335E1"
                     /clone_lib="RPCI-3"
     repeat_region   9. .54
                     /note="L2 repeat: matches 2705. .2750 of consensus"
     repeat_region   546. .638
                     /note="LTR33 repeat: matches 256. .355 of consensus"
     repeat_region   693. .849
                     /note="MER69A repeat: matches 22. .175 of consensus"
     repeat_region   849. .963
                     /note="LTR33 repeat: matches 403. .518 of consensus"
     repeat_region   2404. .2638
                     /note="L2 repeat: matches 2491. .2743 of consensus"
     misc_feature    complement(2642. .3106)
                     /note="match: GSS: Em:AQ629085"
     misc_feature    complement(2688. .3066)
                     /note="match: GSS: Em:B85272"
     repeat_region   2947. .4161
                     /note="L1MA7 repeat: matches 5095. .6282 of consensus"
     repeat_region   4208. .4526
                     /note="AluSx repeat: matches 1. .312 of consensus"
     repeat_region   4528. .4625
                     /note="49 copies 2 mer aa 65% conserved"
```

```
repeat_region   4776. .5058
                /note="AluSq repeat: matches 17. .301 of consensus"
repeat_region   5549. .5858
                /note="AluSx repeat: matches 1. .300 of consensus"
repeat_region   5897. .5996
                /note="MIR repeat: matches 76. .182 of consensus"
repeat_region   9252. .9688
                /note="L1MB2 repeat: matches 5769. .6168 of consensus"
repeat_region   9698. .10507
                /note="L1PA10 repeat: matches 5295. .6164 of consensus"
repeat_region   10676. .10709
                /note="17 copies 2 mer aa 82% conserved"
repeat_region   10805. .10844
                /note="L2 repeat: matches 2671. .2710 of consensus"
misc_feature    complement(11164. .11305)
                /note="match: GSS: Em:AQ070712"
misc_feature    11450. .11804
                /note="match: GSS: Em:AQ253061"
repeat_region   13851. .14130
                /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region   14304. .14351
                /note="L1M3d repeat: matches 251. .300 of consensus"
repeat_region   15500. .16053
                /note="L1M2 repeat: matches 697. .1258 of consensus"
misc_feature    16056. .16439
                /note="match: GSS: Em:AQ123207"
repeat_region   16105. .16169
                /note="L1M2 repeat: matches 1252. .1315 of consensus"
repeat_region   16143. .16306
                /note="L1M2 repeat: matches 1824. .2364 of consensus"
misc_feature    16257. .16436
                /note="match: GSS: Em:AQ508275 Em:AQ750035 Em:B85162"
repeat_region   16505. .17689
                /note="L1M1 repeat: matches 1908. .3130 of consensus"
repeat_region   17686. .18668
                /note="L1MA6 repeat: matches 3638. .4613 of consensus"
repeat_region   18669. .18963
                /note="AluJo repeat: matches 1. .290 of consensus"
repeat_region   18964. .19047
                /note="L1MA6 repeat: matches 4613. .4695 of consensus"
repeat_region   19048. .19378
                /note="AluJo repeat: matches 1. .305 of consensus"
repeat_region   19379. .19801
                /note="L1MA6 repeat: matches 4695. .5139 of consensus"
repeat_region   19780. .20811
                /note="L1MA6 repeat: matches 5285. .6297 of consensus"
repeat_region   20856. .21173
                /note="AluY repeat: matches 1. .311 of consensus"
repeat_region   21177. .21256
                /note="MLT1E repeat: matches 11. .81 of consensus"
repeat_region   21450. .21616
                /note="MLT1E repeat: matches 286. .460 of consensus"
repeat_region   21682. .21815
                /note="LTR33 repeat: matches 374. .521 of consensus"
repeat_region   21914. .22119
                /note="LTR33 repeat: matches 5. .221 of consensus"
repeat_region   22394. .22698
                /note="AluYa5 repeat: matches 1. .304 of consensus"
repeat_region   22698. .22971
                /note="AluSx repeat: matches 7. .281 of consensus"
repeat_region   22984. .23284
                /note="AluY repeat: matches 1. .298 of consensus"
repeat_region   23528. .24208
                /note="L1MB5 repeat: matches 4919. .5626 of consensus"
repeat_region   24212. .24350
                /note="MLT1A2 repeat: matches 4. .143 of consensus"
repeat_region   24346. .24422
                /note="MLT1A2 repeat: matches 294. .374 of consensus"
repeat_region   24423. .24695
                /note="L1MB5 repeat: matches 5622. .5921 of consensus"
repeat_region   24697. .25040
                /note="MER1A repeat: matches 1. .347 of consensus"
repeat_region   26627. .26852
                /note="MIR repeat: matches 5. .262 of consensus"
repeat_region   27852. .28161
                /note="L2 repeat: matches 2074. .2398 of consensus"
repeat_region   29319. .29456
                /note="FRAM repeat: matches 1. .151 of consensus"
repeat_region   30147. .30209
                /note="L1MD3 repeat: matches 6602. .6668 of consensus"
repeat_region   30738. .31054
                /note="AluJb repeat: matches 1. .306 of consensus"
misc_feature    31861. .32423
                /note="match: STS: Em:AA134931"
misc_feature    31989. .32126
                /note="match: STS: Em:G44108"
repeat_region   33086. .33248
                /note="AluSg/x repeat: matches 127. .289 of consensus"
repeat_region   33329. .33490
                /note="AluJb repeat: matches 136. .296 of consensus"
repeat_region   33675. .33782
                /note="L2 repeat: matches 2096. .2209 of consensus"
repeat_region   34176. .34821
                /note="L2 repeat: matches 2056. .2710 of consensus"
repeat_region   34972. .35128
                /note="AluJb repeat: matches 132. .288 of consensus"
repeat_region   35810. .35950
                /note="MER46C repeat: matches 59. .196 of consensus"
repeat_region   36166. .36486
                /note="AluJb repeat: matches 1. .312 of consensus"
repeat_region   38407. .38437
                /note="L2 repeat: matches 2690. .2720 of consensus"
repeat_region   39159. .39454
                /note="AluSq repeat: matches 1. .296 of consensus"
repeat_region   40502. .40795
                /note="AluJo repeat: matches 1. .294 of consensus"
repeat_region   41209. .41505
                /note="AluY repeat: matches 1. .308 of consensus"
repeat_region   42583. .42890
                /note="AluSg repeat: matches 1. .310 of consensus"
repeat_region   43281. .43376
                /note="L1ME repeat: matches 5304. .5395 of consensus"
repeat_region   43377. .43652
                /note="AluY repeat: matches 1. .310 of consensus"
repeat_region   43653. .44046
                /note="L1ME repeat: matches 5395. .5813 of consensus"
repeat_region   45125. .45371
                /note="L2 repeat: matches 2373. .2626 of consensus"
repeat_region   46153. .46351
                /note="MSTC repeat: matches 1. .183 of consensus"
repeat_region   46456. .46593
                /note="MSTC repeat: matches 253. .394 of consensus"
misc_feature    46601. .46944
                /note="match: GSS: Em:AQ097981"
repeat_region   47206. .47513
                /note="AluSx repeat: matches 1. .300 of consensus"
repeat_region   48740. .48838
                /note="HY3 repeat: matches 2. .100 of consensus"
repeat_region   50464. .50617
```

```
  Query Match             70.4%;  Score 17.6;  DB 9;  Length 143805;
  Best Local Similarity   83.3%;  Pred. No. 5.3e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgcag 25
          ||||| || |||| |||||| |||
Db  90879 CATTTGGAGCCTCATCAGGTTCAG 90902


RESULT  40
AC090711
LOCUS       AC090711   153064 bp    DNA             HTG       02-JUN-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-89H1 map 8q21.1, ***
            SEQUENCING IN PROGRESS ***, 1 ordered pieces.
ACCESSION   AC090711
VERSION     AC090711.2  GI:14280158
```

```
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153064)
  AUTHORS   Schilhabel,M.B., Baumgart,C., Blechschmidt,K., Dette,M., Jahn,N.,
            Lehmann,R., Menzel,U., Polley,A., Reichwald,K., Schudy,A.,
            Siddiqui,R., Taudien,S., Wen,G., Rosenthal,A. and Platzer,M.
  TITLE     Chromosome 8 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 153064)
  AUTHORS   Wen,G., Rosenthal,A., Schattevoy,R., Baumgart,C. and Platzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2001) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT     On Jun 2, 2001 this sequence version replaced gi:13259420.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   153064: contig of 153064 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153064
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q21.1"
                     /clone="RP11-89H1"
BASE COUNT    47753 a  27451 c  28431 g  49429 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 2;  Length 153064;
  Best Local Similarity   83.3%;  Pred. No. 5.4e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           ||| |||||||| |||| |||||
Db 121527 CATGTCGATCCTATTCAGTTGCAG 121550


RESULT  41
AC023079/c
LOCUS       AC023079   157254 bp    DNA             HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 14 clone RP11-59P19 map 14, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC023079
VERSION     AC023079.3  GI:8954144
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 157254)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 14, clone RP11-59P19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 157254)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 7, 2000 this sequence version replaced gi:7139783.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6308
                Center clone name: 59_P_19
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Sequencing vector: Plasmid; n/a; %-0.f%% of reads
                0.495399858457183Chemistry: Dye-terminator Big Dye; 100% of
            reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 142719 bases at least Q40
                Consensus quality: 150008 bases at least Q30
                Consensus quality: 152903 bases at least Q20
                Insert size: 109000; agarose-fp
                Insert size: 154754; sum-of-contigs
                Quality coverage: 6.1 in Q20 bases; agarose-fp
                Quality c.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1263: contig of 1263 bp in length
            *    1264 1363: gap of      100 bp
            *    1364     2624: contig of 1261 bp in length
            *    2625 2724: gap of      100 bp
            *    2725     4379: contig of 1655 bp in length
            *    4380 4479: gap of      100 bp
            *    4480     6168: contig of 1689 bp in length
            *    6169 6268: gap of      100 bp
            *    6269     7502: contig of 1234 bp in length
            *    7503 7602: gap of      100 bp
            *    7603     9227: contig of 1625 bp in length
            *    9228 9327: gap of      100 bp
            *    9328    11272: contig of 1945 bp in length
            *   11273 11372: gap of      100 bp
            *   11373    14027: contig of 2655 bp in length
            *   14028 14127: gap of      100 bp
            *   14128    16796: contig of 2669 bp in length
            *   16797 16896: gap of      100 bp
            *   16897    19661: contig of 2765 bp in length
            *   19662 19761: gap of      100 bp
            *   19762    21731: contig of 1970 bp in length
            *   21732 21831: gap of      100 bp
            *   21832    25103: contig of 3272 bp in length
            *   25104 25203: gap of      100 bp
            *   25204    27943: contig of 2740 bp in length
            *   27944 28043: gap of      100 bp
            *   28044    32560: contig of 4517 bp in length
            *   32561 32660: gap of      100 bp
            *   32661    39551: contig of 6891 bp in length
```

```
            *    39552 39651: gap of      100 bp
            *    39652    44837: contig of 5186 bp in length
            *    44838 44937: gap of      100 bp
            *    44938    52684: contig of 7747 bp in length
            *    52685 52784: gap of      100 bp
            *    52785    60087: contig of 7303 bp in length
            *    60088 60187: gap of      100 bp
            *    60188    68029: contig of 7842 bp in length
            *    68030 68129: gap of      100 bp
            *    68130    76334: contig of 8205 bp in length
            *    76335 76434: gap of      100 bp
            *    76435    88231: contig of 11797 bp in length
            *    88232 88331: gap of      100 bp
            *    88332    99697: contig of 11366 bp in length
            *    99698 99797: gap of      100 bp
            *    99798   112971: contig of 13174 bp in length
            *   112972 113071: gap of      100 bp
            *   113072   125118: contig of 12047 bp in length
            *   125119 125218: gap of      100 bp
            *   125219   142067: contig of 16849 bp in length
            *   142068 142167: gap of      100 bp
            *   142168   157254: contig of 15087 bp in length.
FEATURES             Location/Qualifiers
     source          1. .157254
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14"
                     /clone="RP11-59P19"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1263
                     /note="assembly_fragment"
     misc_feature    1364. .2624
                     /note="assembly_fragment"
     misc_feature    2725. .4379
                     /note="assembly_fragment"
     misc_feature    4480. .6168
                     /note="assembly_fragment"
     misc_feature    6269. .7502
                     /note="assembly_fragment"
     misc_feature    7603. .9227
                     /note="assembly_fragment"
     misc_feature    9328. .11272
                     /note="assembly_fragment"
     misc_feature    11373. .14027
                     /note="assembly_fragment"
     misc_feature    14128. .16796
                     /note="assembly_fragment"
     misc_feature    16897. .19661
                     /note="assembly_fragment"
     misc_feature    19762. .21731
                     /note="assembly_fragment"
     misc_feature    21832. .25103
                     /note="assembly_fragment"
     misc_feature    25204. .27943
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    28044. .32560
                     /note="assembly_fragment"
     misc_feature    32661. .39551
                     /note="assembly_fragment"
     misc_feature    39652. .44837
                     /note="assembly_fragment"
     misc_feature    44938. .52684
                     /note="assembly_fragment"
     misc_feature    52785. .60087
                     /note="assembly_fragment"
     misc_feature    60188. .68029
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    68130. .76334
                     /note="assembly_fragment"
     misc_feature    76435. .88231
                     /note="assembly_fragment"
     misc_feature    88332. .99697
                     /note="assembly_fragment"
     misc_feature    99798. .112971
                     /note="assembly_fragment"
     misc_feature    113072. .125118
                     /note="assembly_fragment"
     misc_feature    125219. .142067
                     /note="assembly_fragment"
     misc_feature    142168. .157254
                     /note="assembly_fragment"
BASE COUNT    44228 a  32822 c  33152 g  44549 t   2503 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 2;  Length 157254;
  Best Local Similarity   83.3%;  Pred. No. 5.4e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgcag 25
          ||||| | |||||| ||| |||||
Db 125667 CATTTGGGTCCTCGGCAGCTGCAG 125644


RESULT  42
AC007580
LOCUS       AC007580   160903 bp    DNA             INV       28-FEB-2001
DEFINITION  Drosophila melanogaster, chromosome 2R, region 50F-51C, BAC clone
            BACR07P02, complete sequence.
ACCESSION   AC007580
VERSION     AC007580.5  GI:13162480
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 160903)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2R, region 50F-51C
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 160903)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Feb 28, 2001 this sequence version replaced gi:5670594.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
```

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdgp@fruitfly.berkeley.edu.

```
FEATURES             Location/Qualifiers
     source          1. .160903
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="50F-51C"
                     /clone="BACR07P02 (D623)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    45488 a  33530 c  34013 g  47872 t
ORIGIN
```

```
  Query Match             70.4%;  Score 17.6;  DB 3;  Length 160903;
  Best Local Similarity   83.3%;  Pred. No. 5.4e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           || || |||||| |||||| ||||
Db   13960 CAATTGGATCCTGGTCAGGAGCAG 13983
```

```
RESULT  43
AC008306
LOCUS       AC008306   169689 bp    DNA             INV       21-MAR-2001
DEFINITION  Drosophila melanogaster, chromosome 2R, region 51B-51B, BAC clone
            BACR27M04, complete sequence.
ACCESSION   AC008306
VERSION     AC008306.3  GI:13399345
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 169689)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2R, region 51B-51B
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169689)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Mar 21, 2001 this sequence version replaced gi:5738288.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .169689
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
                     /map="51B-51B"
                     /clone="BACR27M04 (D708)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    48346 a  35214 c  34999 g  51130 t
ORIGIN
```

```
  Query Match             70.4%;  Score 17.6;  DB 3;  Length 169689;
  Best Local Similarity   83.3%;  Pred. No. 5.4e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           || || |||||| |||||| ||||
Db 134076 CAATTGGATCCTGGTCAGGAGCAG 134099
```

```
RESULT  44
AL160251
LOCUS       AL160251   171398 bp    DNA             HTG       26-JUL-2001
DEFINITION  Homo sapiens chromosome 13 clone RP11-102K13, *** SEQUENCING IN
            PROGRESS ***, 4 unordered pieces.
ACCESSION   AL160251
VERSION     AL160251.22  GI:15026933
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171398)
  AUTHORS   Lloyd,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 27, 2001 this sequence version replaced gi:15020861.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA102K13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 0% of reads
            Sequencing vector: plasmid; L08752; 99% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 169835 bases at least Q40
            Consensus quality: 170044 bases at least Q30
            Consensus quality: 170283 bases at least Q20
            Insert size: 171098; sum-of-contigs
            Insert size: 192491; 10.8% error; agarose-fp
            Quality coverage: 13.61x in Q20 bases; sum-of-contigs Quality
            coverage: 12.98x in Q20 bases; agarose-fp
            --------------.
```

```
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   120243: contig of 120243 bp in length
            *  120244 120343: gap of      100 bp
            *  120344   130243: contig of 9900 bp in length
            *  130244 130343: gap of      100 bp
            *  130344   146268: contig of 15925 bp in length
            *  146269 146368: gap of      100 bp
            *  146369   171398: contig of 25030 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171398
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-102K13"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .120243
                     /note="assembly_fragment:03775"
     misc_feature    120344. .130243
                     /note="assembly_fragment:01208
                     fragment_chain:1"
     misc_feature    130344. .146268
                     /note="assembly_fragment:03011
                     fragment_chain:1"
     misc_feature    146369. .171398
                     /note="assembly_fragment:08635
                     fragment_chain:1
                     clone_end:T7
                     vector_side:right"
BASE COUNT    44677 a  42777 c  42228 g  41414 t    302 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 2;  Length 171398;
  Best Local Similarity   83.3%;  Pred. No. 5.4e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgcag 25
          | ||||||| ||| |||| |||||
Db 100491 CCTTTCGATTCTCCTCAGCTGCAG 100514


RESULT  45
HSU52112
LOCUS       HSU52112   174424 bp    DNA             PRI       04-JAN-2001
DEFINITION  Homo sapiens Xq28 genomic DNA in the region of the L1CAM locus
            containing the genes for neural cell adhesion molecule L1 (L1CAM),
            arginine-vasopressin receptor (AVPR2), C1 p115 (C1), ARD1
            N-acetyltransferase related protein (TE2), renin-binding protein
            (RbP), host cell factor 1 (HCF1), and interleukin-1
            receptor-associated kinase (IRAK) genes, complete cds, and Xq28lu2
            gene.
ACCESSION   U52112
VERSION     U52112.1  GI:1302657
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174424)
  AUTHORS   Brenner,V., Nyakatura,G., Rosenthal,A. and Platzer,M.
  TITLE     Genomic organization of two novel genes on human Xq28: compact head
            to head arrangement of IDH gamma and TRAP delta is conserved in rat
            and mouse
  JOURNAL   Genomics 44 (1), 8-14 (1997)
  MEDLINE   97432815
   PUBMED   9286695

REFERENCE   2  (bases 1 to 174424)
  AUTHORS   Platzer,M., Bauer,D. and Drescher,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-1995) Matthias Platzer, Department of Genome
            Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11,
            Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1. .174424
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xq28"
     repeat_region   complement(1. .62)
                     /rpt_family="Alu"
     repeat_region   complement(191. .478)
                     /rpt_family="Alu-Ya"
     repeat_region   complement(547. .833)
                     /rpt_family="Alu-Sp"
     repeat_region   1685. .1979
                     /rpt_family="Alu-Sx"
     repeat_region   complement(2815. .3104)
                     /rpt_family="Alu-Sx"
     repeat_region   complement(3242. .3593)
                     /rpt_family="Alu-Jo"
     repeat_region   complement(4509. .4821)
                     /rpt_family="Alu-Sp"
     repeat_region   5413. .5697
                     /rpt_family="Alu-Jo"
     repeat_region   complement(6126. .6295)
                     /rpt_family="Alu-J"
     repeat_region   complement(7050. .7310)
                     /rpt_family="Alu-Jb"
     repeat_region   complement(7402. .7599)
                     /rpt_family="L1MA2"
     repeat_region   8253. .8529
                     /rpt_family="Alu-Jo"
     repeat_region   complement(9747. .9926)
                     /rpt_family="Alu-Jb"
     repeat_region   10090. .10194
                     /rpt_family="Alu-J"
     repeat_region   11157. .11364
                     /rpt_family="L1MB3"
     repeat_region   11593. .11890
                     /rpt_family="L1MC2"
     repeat_region   12083. .12370
                     /rpt_family="Alu-Sx"
     repeat_region   13165. .13386
                     /rpt_family="Alu-Jo"
     repeat_region   complement(13727. .13784)
                     /rpt_family="Alu-Sxzg"
     repeat_region   complement(13801. .13948)
                     /rpt_family="L1ME2"
     repeat_region   13951. .14005
                     /rpt_family="Alu-J"
     3'UTR           complement(14868. .15598)
                     /gene="L1CAM"
     mRNA            complement(join(14868. .15830,16304. .16315,16413. .16485,
                     16819. .16953,17258. .17413,17521. .17640,17757. .17930,
                     18024. .18146,18235. .18436,18640. .18755,19585. .19807,
                     19990. .20060,20291. .20488,20735. .20845,20933. .21057,
                     21237. .21393,21496. .21662,21776. .21887,22455. .22598,
                     22738. .22869,22991. .23175,23323. .23434,23725. .23895,
                     23992. .24114,25090. .25292,25530. .25635,26166. .26180,
                     28699. .28793))
                     /gene="L1CAM"
     gene            complement(14868. .28793)
                     /gene="L1CAM"
     exon            complement(14868. .15830)
                     /gene="L1CAM"
                     /number=28
     CDS             complement(join(15599. .15830,16304. .16315,16413. .16485,
                     16819. .16953,17258. .17413,17521. .17640,17757. .17930,
                     18024. .18146,18235. .18436,18640. .18755,19585. .19807,
```

```
                19990. .20060,20291. .20488,20735. .20845,20933. .21057,
                21237. .21393,21496. .21662,21776. .21887,22455. .22598,
                22738. .22869,22991. .23175,23323. .23434,23725. .23895,
                23992. .24114,25090. .25292,25530. .25635,26166. .26180,
                28699. .28774))
                /gene="L1CAM"
                /codon_start=1
                /product="neural cell adhesion molecule L1"
                /protein_id="AAC51746.1"
                /db_xref="GI:1302658"
                /translation="MVVALRYVWPLLLCSPCLLIQIPEEYEGHHVMEPPVITEQSPRR
                LVVFPTDDISLKCEASGKPEVQFRWTRDGVHFKPKEELGVTVYQSPHSGSFTITGNNS
                NFAQRFQGIYRCFASNKLGTAMSHEIRLMAEGAPKWPKETVKPVEVEEGESVVLPCNP
                PPSAEPLRIYWMNSKILHIKQDERVTMGQNGNLYFANVLTSDNHSDYICHAHFPGTRT
                IIQKEPIDLRVKATNSMIDRKPRLLFPTNSSSHLVALQGQPLVLECIAEGFPTPTIKW
                LRPSGPMPADRVTYQNHNKTLQLLKVGEEDDGEYRCLAENSLGSARHAYYVTVEAAPY
                WLHKPQSHLYGPGETARLDCQVQGRPQPEVTWRINGIPVEELAKDQKYRIQRGALILS
                NVQPSDTMVTQCEARNRHGLLLANAYIYVVQLPAKILTADNQTYMAVQGSTAYLLCKA
                FGAPVPSVQWLDEDGTTVLQDERFFPYANGTLGIRDLQANDTGRYFCLAANDQNNVTI
                MANLKVKDATQITQGPRSTIEKKGSRVTFTCQASFDPSLQPSITWRGDGRDLQELGDS
                DKYFIEDGRLVIHSLDYSDQGNYSCVASTELDVVESRAQLLVVGSPGPVPRLVLSDLH
                LLTQSQVRVSWSPAEDHNAPIEKYDIEFEDKEMAPEKWYSLGKVPGNQTSTTLKLSPY
                VHYTFRVTAINKYGPGEPSPVSETVVTPEAAPEKNPVDVKGEGNETTNMVITWKPLRW
                MDWNAPQVQYRVQWRPQGTRGPWQEQIVSDPFLVVSNTSTFVPYEIKVQAVNSQGKGP
                EPQVTIGYSGEDYPQAIPELEGIEILNSSAVLVKWRPVDLAQVKGHLRGYNVTYWREG
                SQRKHSKRHIHKDHVVVPANTTSVILSGLRPYSSYHLEVQAFNGRGSGPASEFTFSTP
                EGVPGHPEALHLECQSNTSLLLRWQPPLSHNGVLTGYVLSYHPLDEGGKGQLSFNLRD
                PELRTHNLTDLSPHLRYRFQLQATTKEGPGEAIVREGGTMALSGISDFGNISATAGEN
                YSVVSWVPKEGQCNFRFHILFKALGEEKGGASLSPQYVSYNQSSYTQWDLQPDTDYEI
                HLFKERMFRHQMAVKTNGTGRVRLPPAGFATEGWFIGFVSAIILLLLVLLILCFIKRS
                KGGKYSVKDKEDTQVDSEARPMKDETFGEYRSLESDNEEKAFGSSQPSLNGDIKPLGS
                DDSLADYGGSVDVQFNEDGSFIGQYSGKKEKEAAGGNDSSGATSPINPAVALE"
exon            complement(16304. .16315)
                /gene="L1CAM"
                /number=27
exon            complement(16413. .16485)
                /gene="L1CAM"
                /number=26
exon            complement(16819. .16953)
                /gene="L1CAM"
                /number=25
exon            complement(17258. .17413)
                /gene="L1CAM"
                /number=24
exon            complement(17521. .17640)
                /gene="L1CAM"
                /number=23
exon            complement(17757. .17930)
                /gene="L1CAM"
                /number=22
exon            complement(18024. .18146)
                /gene="L1CAM"
                /number=21
exon            complement(18235. .18436)
                /gene="L1CAM"
                /number=20
exon            complement(18640. .18755)
                /gene="L1CAM"
                /number=19
repeat_region   18969. .19256
                /rpt_family="Alu-Sz"
repeat_region   19418. .19501
                /rpt_family="MER42C"
exon            complement(19585. .19807)
                /gene="L1CAM"
                /number=18
exon            complement(19990. .20060)
                /gene="L1CAM"
                /number=17
exon            complement(20291. .20488)
                /gene="L1CAM"
                /number=16
exon            complement(20735. .20845)
                /gene="L1CAM"
                /number=15
exon            complement(20933. .21057)
                /gene="L1CAM"
                /number=14
exon            complement(21237. .21393)
                /gene="L1CAM"
                /number=13
exon            complement(21496. .21662)
                /gene="L1CAM"
                /number=12
exon            complement(21776. .21887)
                /gene="L1CAM"
                /number=11
exon            complement(22455. .22598)
                /gene="L1CAM"
                /number=10
exon            complement(22738. .22869)
                /gene="L1CAM"
                /number=9
exon            complement(22991. .23175)
                /gene="L1CAM"
                /number=8
exon            complement(23323. .23434)
                /gene="L1CAM"
                /number=7
exon            complement(23725. .23895)
                /gene="L1CAM"
                /number=6
exon            complement(23992. .24114)
                /gene="L1CAM"

  Query Match             70.4%;  Score 17.6;  DB 9;  Length 174424;
  Best Local Similarity   83.3%;  Pred. No. 5.4e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           |||||| || | ||||||| ||||
Db   11360 CATTTCCATACCCGTCAGGAGCAG 11383


Search completed: December 26, 2001, 10:58:38
Job time: 2080 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:19:25 ; Search time 366.42 Seconds
(without alignments)
58.493 Million cell updates/sec

Title: US-09-396-196F-8
Perfect score: 25
Sequence: 1 gcatttcgatcctcgtcaggtgcag 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	839	22	AAI62941	Human genomic DNA
	2	25	100.0	1041	20	AAX01303	E. coli biotin syn
	3	25	100.0	1084	10	AAN91329	E.coli Bio B gene.
	4	25	100.0	1121	7	AAN60496	Sequence encoding
	5	25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
	6	18.2	72.8	127197	22	AAI61370	Soybean 515002 req
	7	17.6	70.4	295	22	AAH70365	Human cervical can
	8	17.6	70.4	295	22	AAH73070	Human cervical can
	9	17.6	70.4	321	22	AAH69625	Human cervical can
	10	17.6	70.4	565	21	AAC74619	Human ORFX ORF174
c	11	17.2	68.8	241	21	AAZ36290	Mechanical stress
c	12	17.2	68.8	393	20	AAV80637	Kidney injury asso
	13	17	68.0	1542	22	AAC82667	Human 7TM clone HN
	14	17	68.0	1782	22	AAF33079	Human secreted pro
c	15	17	68.0	2240	22	AAD06664	A. thaliana transc
c	16	17	68.0	2240	22	AAD05746	Arabidopsis thalia
c	17	17	68.0	2240	22	AAD05773	Arabidopsis thalia
c	18	17	68.0	2254	19	AAV65382	Arabidopsis phytoc
c	19	17	68.0	3234	12	AAQ15019	Encodes mouse bull
c	20	16.6	66.4	34094	20	AAZ30163	Complete nucleotid
	21	16.4	65.6	1356	22	AAH67831	C glutamicum codin
c	22	16.4	65.6	349980	22	AAH68533	C glutamicum codin
c	23	16.2	64.8	385	21	AAC56248	Pinus radiata tran
c	24	16.2	64.8	393	21	AAC57313	Pinus radiata tran
	25	16.2	64.8	865	18	AAX30823	Streptococcus pneu
c	26	16.2	64.8	882	19	AAT98665	DNA encoding a S.
	27	16.2	64.8	1702	19	AAZ96336	S. pneumoniae deri
	28	16.2	64.8	1702	19	AAV37384	Streptococcus pneu
c	29	16.2	64.8	14736	19	AAV52304	Streptococcus pneu
c	30	16	64.0	601	22	AAH28893	Drosophila melanog
c	31	16	64.0	1045	18	AAT59977	5' untranslated re
c	32	16	64.0	1045	20	AAX02016	D. melanogaster ti
c	33	16	64.0	1137	20	AAX36567	Human EDG-3 coding
c	34	16	64.0	1137	21	AAZ22300	Human EDG3sb polyp
c	35	16	64.0	1137	22	AAS00258	LPA receptor-relat
c	36	16	64.0	1140	22	AAF84497	Rat G protein-coup
c	37	16	64.0	1152	21	AAA65913	E. coli proliferat
c	38	16	64.0	1152	21	AAA39397	Escherichia coli m
c	39	16	64.0	1792	20	AAX59380	Aspergillus nidula
	40	16	64.0	2489	22	AAI59340	Human polynucleoti
	41	16	64.0	2567	20	AAZ52889	Human prostate tum
c	42	16	64.0	3708	13	AAQ23963	Probe BtY2. Bos t
c	43	16	64.0	3904	21	AAA53179	Streptococcus bovi
c	44	16	64.0	3954	18	AAT59975	Full length tipE p
c	45	16	64.0	3954	20	AAX02015	D. melanogaster ti
c	46	16	64.0	4928	16	AAQ94735	Lycopene cyclase g
c	47	16	64.0	4928	17	AAT40205	Cyanobacterial lyc
c	48	16	64.0	11854	20	AAX13243	Enterococcus faeca
c	49	16	64.0	20757	20	AAX20599	Polynucleotide seq
	50	16	64.0	143068	21	AAF21105	Human low adenosin
	51	16	64.0	143068	21	AAF21272	Human low adenosin
	52	16	64.0	143068	21	AAA34983	Human adenosine re
	53	16	64.0	143068	21	AAA35150	Human adenosine re
	54	16	64.0	149412	21	AAA35151	Human adenosine re
	55	16	64.0	152740	21	AAF21273	Human low adenosin
c	56	16	64.0	534720	19	AAV30458	Rhizobium species
c	57	16	64.0	536165	19	AAV30459	Rhizobium species
	58	15.8	63.2	220	22	AAF28005	Human TR13 coding
	59	15.8	63.2	220	22	AAF28044	Human TR13 coding
	60	15.8	63.2	625	21	AAC98199	Human colon cancer
c	61	15.8	63.2	1285	21	AAC57345	Pinus radiata tran
	62	15.8	63.2	1717	20	AAZ41991	Human endometrium
	63	15.8	63.2	2554	22	AAF27997	Human TR13 recepto
	64	15.8	63.2	2610	22	AAF89775	Nucleotide sequenc
	65	15.8	63.2	2733	22	AAF89774	Nucleotide sequenc
	66	15.8	63.2	3280	21	AAA95442	Human CASB619 prot
	67	15.8	63.2	3331	22	AAF89765	Nucleotide sequenc
	68	15.8	63.2	3334	22	AAF28030	Human TR13 recepto
	69	15.8	63.2	3501	22	AAS21262	Human cDNA sequenc
c	70	15.8	63.2	3568	21	AAZ99491	DNA encoding a mai
c	71	15.8	63.2	3776	21	AAZ58263	Corn cellulose syn
c	72	15.8	63.2	5269	21	AAC76615	Human ORFX ORF2170
	73	15.6	62.4	349	22	AAH97983	Murine 7-transmemb
c	74	15.6	62.4	377	21	AAC57197	Eucalyptus grandis
	75	15.6	62.4	522	22	AAH97981	Murine 7-transmemb
	76	15.6	62.4	545	22	AAH97976	Murine 7-transmemb
	77	15.6	62.4	545	22	AAH97978	Murine 7-transmemb
	78	15.6	62.4	550	22	AAH97977	Murine 7-transmemb
	79	15.6	62.4	561	22	AAH97975	Murine 7-transmemb
	80	15.6	62.4	569	22	AAH97974	Murine 7-transmemb
	81	15.6	62.4	575	22	AAH97980	Murine 7-transmemb
	82	15.6	62.4	585	22	AAH97979	Murine 7-transmemb
	83	15.6	62.4	682	21	AAF12087	Aspergillus oryzae
	84	15.6	62.4	700	22	AAH93217	Human inflammatory

85	15.6	62.4	700	22	AAH93218	Human inflammatory
86	15.6	62.4	735	20	AAX91722	Porphorymonas ging
c 87	15.6	62.4	872	22	AAF97936	Human secreted pro
88	15.6	62.4	879	20	AAX91595	Porphorymonas ging
c 89	15.6	62.4	996	17	AAT59342	c-mpl ligand (1-33
c 90	15.6	62.4	1084	17	AAT34429	Eucalyptus SQE1 cD
c 91	15.6	62.4	1136	21	AAZ52531	Human secreted pro
c 92	15.6	62.4	1167	20	AAZ24889	Human secreted pro
c 93	15.6	62.4	1167	22	AAF97938	Human secreted pro
c 94	15.6	62.4	1189	22	AAF97904	Human secreted pro
c 95	15.6	62.4	1208	22	AAF97937	Human secreted pro
c 96	15.6	62.4	1223	21	AAC55901	Eucalyptus grandis
c 97	15.6	62.4	1248	21	AAC33385	Arabidopsis thalia
98	15.6	62.4	1438	21	AAC41115	Arabidopsis thalia
99	15.6	62.4	1472	21	AAC49009	Arabidopsis thalia
100	15.6	62.4	1477	21	AAC41264	Arabidopsis thalia

ALIGNMENTS

```
RESULT   1
AAI62941
ID   AAI62941 standard; DNA; 839 BP.
XX
AC   AAI62941;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human genomic DNA SEQ ID NO 269.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant;  immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200155449-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01346.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUL-2000; 2000US-0216880.
PR   14-JUL-2000; 2000US-0218290.
PR   14-AUG-2000; 2000US-0225447.
PR   01-SEP-2000; 2000US-0229343.
PR   06-SEP-2000; 2000US-0230437.
PR   08-SEP-2000; 2000US-0231243.
PR   25-SEP-2000; 2000US-0234997.
PR   29-SEP-2000; 2000US-0236367.
PR   13-OCT-2000; 2000US-0239937.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246528.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249265.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-476225/51.
XX
PT   Novel plasma membrane associated proteins useful for diagnosing,
PT   treating, preventing and/or prognosing disorders related to the
PT   proteins, including cancer, immune response and neuronal disorders  -
XX
PS   Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (AAI62752-AAI62961) and proteins
CC   (AAM42347-AAM42415) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;


  Query Match             100.0%;  Score 25;  DB 22;  Length 839;
  Best Local Similarity   100.0%;  Pred. No. 0.0099;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||||||||||||||||||||||||
Db     131 gcatttcgatcctcgtcaggtgcag 155


RESULT   2
AAX01303
ID   AAX01303 standard; DNA; 1041 BP.
XX
AC   AAX01303;
XX
DT   12-APR-1999  (first entry)
XX
DE   E. coli biotin synthetase (BioB) coding sequence.
XX
KW   DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW   biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS   Escherichia coli.
XX
PN   US5869719-A.
XX
PD   09-FEB-1999.
XX
PF   30-APR-1997;   97US-0846338.
XX
PR   30-APR-1997;   97US-0846338.
PR   08-MAR-1995;   95US-0401068.
XX
```

```
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Patton DA;
XX
DR   WPI; 1999-152902/13.
DR   P-PSDB; AAW73906.
XX
PT   Transgenic plants with high biotin levels - transformed with DNA
PT   encoding di:amino-pelargonic acid amino-transferase or biotin
PT   synthase
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.01;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||||||||||||||||||||||||
Db      99 gcatttcgatcctcgtcaggtgcag 123


RESULT   3
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA;
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 0.01;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||||||||||||||||||||||||
Db     122 gcatttcgatcctcgtcaggtgcag 146


RESULT   4
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzyme -
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 0.01;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||||||||||||||||||||||||
Db     140 gcatttcgatcctcgtcaggtgcag 164


RESULT   5
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
```

```
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry)
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag=  a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag=  b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag=  c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag=  d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag=  e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag=  f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag=  g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag=  h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag=  i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag=  j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag=  k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag=  l
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag=  m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag=  n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                   7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag=  o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag=  p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag=  q
FT                   /standard_name= "rho-independent transcriptional
FT                                   terminator"
FT   stem_loop       5583..5605
FT                   /*tag=  r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;   93WO-EP02688.
XX
PR   02-OCT-1992;   92CH-0003124.
PR   15-JUL-1993;   93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||||||||||||||||||||||||
Db     215 gcatttcgatcctcgtcaggtgcag 239


RESULT   6
AAI61370
ID   AAI61370 standard; DNA; 127197 BP.
XX
AC   AAI61370;
```

```
XX
DT   16-OCT-2001  (first entry)
XX
DE   Soybean 515002 region G2, SEQ ID NO: 1.
XX
KW   Soybean; antihelmintic; gene therapy; soybean cyst nematode; SCN;
KW   SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW   240017 region G3; 318013 region A3; 515002 region G2; ds.
XX
OS   Glycine max.
XX
PN   WO200151627-A2.
XX
PD   19-JUL-2001.
XX
PF   05-JAN-2001; 2001WO-US00552.
XX
PR   07-JAN-2000; 2000US-0174880.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Hauge BM,  Wang ML,  Parsons JD,  Parnell LD;
XX
DR   WPI; 2001-425872/45.
XX
PT   New purified nucleic acid for producing a soybean plant having soybean
PT   cyst nematode resistance and for use in plant breeding programs -
XX
PS   Disclosure; Page 131-204; 1353pp; English.
XX
CC   The invention relates to nucleic acid molecules from regions of the
CC   soybean genome which are associated with soybean cyst nematode (SCN)
CC   resistance. The nucleic acids are used to transform plants, and can
CC   produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC   The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC   of soybean plants and for introgressing SCN resistance or partial SCN
CC   resistance into soybean plants. They can also be used in plant breeding
CC   programmes. The invention also relates to proteins encoded by such
CC   nucleic acid molecules, as well as antibodies capable of recognising
CC   these proteins. The present sequence is a nucleic acid molecule
CC   provided in the specification.
XX
SQ   Sequence 127197 BP; 43836 A; 20975 C; 20437 G; 41949 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 22;  Length 127197;
  Best Local Similarity   87.0%;  Pred. No. 36;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgca 24
           |||| | ||| ||||||||||||
Db   35653 cattgctatcttcgtcaggtgca 35675


RESULT   7
AAH70365
ID   AAH70365 standard; cDNA; 295 BP.
XX
AC   AAH70365;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 1639.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US33312.
XX
PR   08-DEC-1999;   99US-0169681.
PR   21-DEC-1999;   99US-0171350.
PR   14-MAR-2000; 2000US-0189315.
PR   12-MAY-2000; 2000US-0203791.
PR   09-JUN-2000; 2000US-0210600.
PR   21-JUL-2000; 2000US-0220114.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Deeds J,  Berger A,  Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer
PT   and for assessing and detecting compounds for treating the cancer -
XX
PS   Claim 1; Page 363; 1051pp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy.
XX
SQ   Sequence 295 BP; 82 A; 62 C; 58 G; 93 T; 0 other;


  Query Match             70.4%;  Score 17.6;  DB 22;  Length 295;
  Best Local Similarity   83.3%;  Pred. No. 31;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           ||||| || |||| ||||||| |||
Db     132 catttggagcctcatcaggttcag 155


RESULT   8
AAH73070
ID   AAH73070 standard; cDNA; 295 BP.
XX
AC   AAH73070;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 4344.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US33312.
XX
PR   08-DEC-1999;   99US-0169681.
PR   21-DEC-1999;   99US-0171350.
PR   14-MAR-2000; 2000US-0189315.
PR   12-MAY-2000; 2000US-0203791.
PR   09-JUN-2000; 2000US-0210600.
PR   21-JUL-2000; 2000US-0220114.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Deeds J,  Berger A,  Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
```

```
PT   New isolated nucleic acid for diagnosing and treating cervical cancer
PT   and for assessing and detecting compounds for treating the cancer -
XX
PS   Claim 1; Page 955-956; 1051pp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy.
XX
SQ   Sequence 295 BP; 82 A; 62 C; 58 G; 93 T; 0 other;


  Query Match             70.4%;  Score 17.6;  DB 22;  Length 295;
  Best Local Similarity   83.3%;  Pred. No. 31;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgcag 25
          ||||| || |||| ||||||| |||
Db    132 catttggagcctcatcaggttcag 155


RESULT   9
AAH69625
ID   AAH69625 standard; cDNA; 321 BP.
XX
AC   AAH69625;
XX
DT   19-SEP-2001  (first entry)
XX
DE   Human cervical cancer marker nucleic acid 899.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US33312.
XX
PR   08-DEC-1999;   99US-0169681.
PR   21-DEC-1999;   99US-0171350.
PR   14-MAR-2000; 2000US-0189315.
PR   12-MAY-2000; 2000US-0203791.
PR   09-JUN-2000; 2000US-0210600.
PR   21-JUL-2000; 2000US-0220114.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Deeds J,  Berger A,  Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer
PT   and for assessing and detecting compounds for treating the cancer -
XX
PS   Claim 1; Page 254; 1051pp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy.
XX
SQ   Sequence 321 BP; 83 A; 73 C; 68 G; 97 T; 0 other;


  Query Match             70.4%;  Score 17.6;  DB 22;  Length 321;
  Best Local Similarity   83.3%;  Pred. No. 32;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgcag 25
          ||||| || |||| ||||||| |||
Db    159 catttggagcctcatcaggttcag 182


RESULT  10
AAC74619
ID   AAC74619 standard; cDNA; 565 BP.
XX
AC   AAC74619;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF174 polynucleotide sequence SEQ ID NO:347.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB40410.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 610; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
```

```
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 565 BP; 94 A; 168 C; 197 G; 106 T; 0 other;


  Query Match             70.4%;  Score 17.6;  DB 21;  Length 565;
  Best Local Similarity   83.3%;  Pred. No. 34;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgca 24
           |||  |||| || ||||||||||
Db     390 gcacgtcgaaccgcgtcaggtgca 413


RESULT  11
AAZ36290/c
ID   AAZ36290 standard; cDNA; 241 BP.
XX
AC   AAZ36290;
XX
DT   22-FEB-2000  (first entry)
XX
DE   Mechanical stress induced globulin-related cDNA.
XX
KW   Expressed sequence tag; EST; mechanical stress; gene therapy;
KW   osteoporosis; bone density; bone development; ss.
XX
OS   Rattus sp.
XX
PN   WO9960164-A1.
XX
PD   25-NOV-1999.
XX
PF   14-MAY-1999;   99WO-US11066.
XX
PR   15-MAY-1998;   98US-0085673.
XX
PA   (QUAR-) QUARK BIOTECH INC.
XX
PI   Einat P,  Mor O,  Skaliter R,  Feinstein E,  Faerman A;
XX
DR   WPI; 2000-053304/04.
XX
PT   Identification of stress induced genes for determining risk and
PT   preventing, treating or controlling osteoporosis  -
XX
PS   Claim 25; Fig 2A-1; 308pp; English.
XX
CC   AAZ36290-Z36320 represent genes or expressed sequence tags (ESTs)
CC   identified by the method of the invention. The sequences were isolated
CC   from rat osteoblasts. The specification describes a method for the
CC   identification of genes responsive to a specific mechanical stress. The
CC   method comprises applying the mechanical stress to an organism (tissue
CC   or cells comprising bone cells), isolating the specific cellular
CC   fractions and extracting mRNA from them, and differentially analysing the
CC   mRNA in comparison with control samples. The method is used to identify
CC   genes whose expression is responsive to a specific stress. The identified
CC   genes are employed in determining risk associated with a physiological
CC   or disease state. The risk determination methods are used for testing a
CC   medicament for gene therapy. These medicaments, or genes identified by
CC   the method of the invention, are used for treating, preventing or
CC   controlling a physiological or disease state (especially osteoporosis
CC   or bone density or other factors causing or contributing to osteoporosis
CC   or its symptoms or other conditions involved in mechanical stress or its
CC   lack. The methods can also be used for advancing research or studies in
CC   bone development.
XX
SQ   Sequence 241 BP; 51 A; 53 C; 69 G; 68 T; 0 other;


  Query Match             68.8%;  Score 17.2;  DB 21;  Length 241;
  Best Local Similarity   86.4%;  Pred. No. 47;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgc 23
           || ||| |||||||||||| ||
Db     198 CACTTCCATCCTCGTCAGGGGC 177


RESULT  12
AAV80637/c
ID   AAV80637 standard; cDNA; 393 BP.
XX
AC   AAV80637;
XX
DT   01-MAR-1999  (first entry)
XX
DE   Kidney injury associated molecule SAC_23801 cDNA clone.
XX
KW   Kidney injury associated molecule; kidney injury related molecule;
KW   KIM; tissue growth promotion; regeneration; renal condition;
KW   acute renal failure; acute nephritis; tumour; ds.
XX
OS   Rattus sp.
XX
PN   WO9853071-A1.
XX
PD   26-NOV-1998.
XX
PF   22-MAY-1998;   98WO-US10547.
XX
PR   23-MAY-1997;   97US-0047491.
PR   23-MAY-1997;   97US-0047490.
XX
PA   (BIOJ ) BIOGEN INC.
XX
PI   Cate RL,  Hession CA,  Sanicola-Nadel M,  Wei H;
XX
DR   WPI; 1999-045312/04.
XX
PT   Kidney injury-associated molecule, KIM, polypeptides - upregulated
PT   in injured or regenerating tissues, useful to promote tissue growth
PT   and regeneration, especially to treat renal conditions
XX
PS   Claim 28; Page 191-192; 213pp; English.
XX
CC   The present sequence represents a kidney injury associated molecule
CC   (KIM) cDNA clone. KIM proteins can be administered therapeutically
CC   by expressing KIM encoding polynucleotides, to promote growth and/or
CC   survival of damaged tissue (e.g. renal tissue), since the KIM proteins
CC   are upregulated in injured or regenerating (especially renal) tissues.
CC   KIM fusion proteins, conjugates, antibodies and vectors can also be used
CC   therapeutically, e.g. these or the KIM proteins may be included with an
CC   acceptable carrier in pharmaceutical compositions, useful for therapy/
CC   prophylaxis of conditions associated with disfunction/disregulation of
CC   KIM genes or proteins, especially renal diseases or impairments of renal
CC   function in humans (e.g. acute renal failure, acute nephritis). The
CC   polynucleotides can be used to produce antisense sequences which, when
CC   internalised into cells, can disrupt expression of a cellular KIM gene,
CC   also useful in therapy (e.g. to block the growth of tumours dependent on
CC   KIM for growth) or compositions. The proteins and polynucleotides are
CC   useful diagnostically e.g. to detect and quantify renal injury/disease
CC   (indicative of increased risk, or presence of, renal injury or impaired
CC   function), or abnormal responses to tissue injury (indicative of
CC   increased risk, or presence of, an autoimmune response or abnormal
CC   tissue growth arising from/affecting renal tissue). The proteins can
```

```
CC   also be used to locate KIM-producing cells (especially specific loci,
CC   e.g. tissue masses abnormally producing/expressing KIM such as tumours
CC   arising from/affecting renal tissue), by contacting cells with an
CC   imageable KIM-binding reagent and imaging reagent accumulation.
XX
SQ   Sequence 393 BP; 76 A; 108 C; 112 G; 92 T; 5 other;


  Query Match             68.8%;  Score 17.2;  DB 20;  Length 393;
  Best Local Similarity   86.4%;  Pred. No. 50;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgc 23
           || ||| |||||||||||| ||
Db     258 CACTTCCATCCTCGTCAGGGGC 237


RESULT  13
AAC82667
ID   AAC82667 standard; cDNA; 1542 BP.
XX
AC   AAC82667;
XX
DT   15-MAR-2001  (first entry)
XX
DE   Human 7TM clone HNHEN70 cDNA.
XX
KW   Transmembrane receptor protein; 7TM; gene therapy; human; bactericidal;
KW   fungicidal; viricidal; anti-HIV; analgesic; cytostatic; anabolic;
KW   anti-asthmatic; anti-parkinsonian; cardiant; hypotensive; hypertensive;
KW   osteopathic; anti-ulcer; immunosuppressive; cerebroprotective; vaccine;
KW   antidepressant; neuroprotective; heptahelical receptor; infection;
KW   serpentine receptor; G-protein coupled receptor; cancer; anorexia;
KW   bulimia; asthma; Parkinson's disease; acute heart failure; hypotension;
KW   hypertension; urinary retention; osteoporosis; angina pectoris; ulcer;
KW   myocardial infarction; allergy; benign prostatic hypertrophy; anxiety;
KW   schizophrenia; manic depression; Huntington's chorea; ss.
XX
OS   Homo sapiens.
XX
PN   WO200071584-A1.
XX
PD   30-NOV-2000.
XX
PF   19-MAY-2000; 2000WO-US13737.
XX
PR   20-MAY-1999;   99US-0135167.
PR   13-JUL-1999;   99US-0143616.
PR   09-SEP-1999;   99US-0152934.
PR   14-MAR-2000;  2000US-0189029.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Soppet DR,  Li Y,  Fan P;
XX
DR   WPI; 2001-025139/03.
DR   P-PSDB; AAB45699, AAB45706, AAB45707, AAB45708, AAB45709.
XX
PT   Nucleic acids encoding human 7 transmembrane receptor polypeptides,
PT   useful for preventing, diagnosing and treating e.g. asthma, Parkinson's
PT   disease and ulcers -
XX
PS   Claim 1a; Page 270; 288pp; English.
XX
CC   This invention describes novel nucleic acid molecules (I) encoding human
CC   7 transmembrane (7TM) receptor polypeptides (also called heptahelical,
CC   serpentine or G-protein coupled receptors). The products of the invention
CC   have bactericidal, fungicidal, viricidal, anti-HIV, analgesic,
CC   cytostatic, anabolic, anti-asthmatic, anti-parkinsonian, cardiant,
CC   hypotensive, hypertensive, osteopathic, anti-ulcer, immunosuppressive,
CC   cerebroprotective, antidepressant and neuroprotective activity and can
CC   be used in gene therapy or in a vaccine. (I) and the 7TM receptor protein
CC   it encodes may be used in the prevention, treatment and diagnosis of
CC   diseases associated with inappropriate 7TM receptor expression. (I) and
CC   complementary sequences may also be used as DNA probes in diagnostic
CC   assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate
CC   the presence of similar nucleic acid sequences in samples, and hence
CC   which patients may be in need of restorative therapy. Diseases associated
CC   with inappropriate 7TM receptor polypeptide expression include for
CC   example bacterial, fungal and viral (especially human immuno-deficiency
CC   virus) infections, pain, cancers, anorexia, bulimia, asthma, Parkinson's
CC   disease, acute heart failure, hypotension, hypertension, urinary
CC   retention, osteoporosis, angina pectoris, myocardial infarction, ulcers,
CC   allergies, benign prostatic hypertrophy, anxiety, schizophrenia, manic
CC   depression and Huntington's chorea.
XX
SQ   Sequence 1542 BP; 451 A; 325 C; 306 G; 460 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 22;  Length 1542;
  Best Local Similarity   80.0%;  Pred. No. 75;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||| ||  |||||  |||||||||
Db    1451 gcatctctttcctcaccaggtgcag 1475


RESULT  14
AAF33079
ID   AAF33079 standard; cDNA; 1782 BP.
XX
AC   AAF33079;
XX
DT   23-MAR-2001  (first entry)
XX
DE   Human secreted protein gene 43 SEQ ID NO:53.
XX
KW   Human; secreted protein; diagnosis; cytostatic; antirheumatic;
KW   antiarthritic; dermalogical; cardiant; antiinflammatory; anti-ulcer;
KW   gastrointestinal; solid tumour; rheumatoid arthritis; psoriasis;
KW   diabetic retinopathy; myocardial angiogenesis; Crohn's disease;
KW   ulcer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200077237-A1.
XX
PD   21-DEC-2000.
XX
PF   01-JUN-2000; 2000WO-US14928.
XX
PR   11-JUN-1999;   99US-0138633.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis GA;
XX
DR   WPI; 2001-071280/08.
DR   P-PSDB; AAB64708.
XX
PT   Nucleic acids encoding 49 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating diseases such as tumors,
PT   rheumatoid arthritis, psoriasis and diabetic retinopathy -
XX
PS   Claim 1; Page 455; 520pp; English.
XX
CC   The polynucleotide sequences given in AAF33037 to AAF33085 encode the
CC   human secreted proteins given in AAB64666 to AAB64714. AAB64715 to
CC   AAB64771 represent human secreted polypeptide sequences and proteins
CC   homologous to them, which are given in the exemplification of the present
CC   invention. Human secreted proteins have activities based on the tissues
CC   and cells the genes are expressed in. Examples of activities include:
CC   cytostatic; antirheumatic; antiarthritic; dermalogical; cardiant;
CC   antiinflammatory; gastrointestinal; and anti-ulcer. The polynucleotides
```

```
CC   and polypeptides can be used in the prevention, treatment and diagnosis
CC   of diseases associated with inappropriate polypeptide expression.
CC   Disorders that may be treated or prevented include solid tumours,
CC   rheumatoid arthritis, psoriasis, diabetic retinopathy, myocardial
CC   angiogenesis, Crohn's disease and ulcers. The polynucleotides and their
CC   complementary sequences may also be used as DNA probes in diagnostic
CC   assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate
CC   the presence of similar nucleic acid sequences in samples, and therefore
CC   which patients may be in need of restorative therapy. The polypeptides
CC   may also be used as antigens in the production of antibodies against the
CC   polypeptide and in assays to identify modulators (agonists and
CC   antagonists) of polypeptide expression and activity. The anti-polypeptide
CC   antibodies and antagonists may also be used to down regulate expression
CC   and activity. AAF33028 to AAF33036 and AAB64665 represent sequences used
CC   in the exemplification of the present invention.
XX
SQ   Sequence 1782 BP; 500 A; 382 C; 353 G; 547 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 22;  Length 1782;
  Best Local Similarity   80.0%;  Pred. No. 77;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||| ||  |||||  |||||||||
Db   1691 gcatctctttcctcaccaggtgcag 1715


RESULT  15
AAD06664/c
ID   AAD06664 standard; cDNA; 2240 BP.
XX
AC   AAD06664;
XX
DT   10-AUG-2001  (first entry)
XX
DE   A. thaliana transcription factor G680 homolog, G214 cDNA.
XX
KW   Plant transcription factor; phenotype; sugar sensing characteristic;
KW   transgenic plant; plant yield; growth; germination; photosynthesis;
KW   glyoxylate metabolism; respiration; pathogen response; wounding response;
KW   cell cycle regulation; pigmentation; flowering; senescence; physiology;
KW   storage organ; metabolism; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             238..2064
FT                   /*tag=  a
FT                   /product= "Transcription factor homolog"
XX
PN   WO200135725-A1.
XX
PD   25-MAY-2001.
XX
PF   14-NOV-2000; 2000WO-US31414.
XX
PR   17-NOV-1999;   99US-0166228.
PR   17-APR-2000; 2000US-0197899.
PR   22-AUG-2000; 2000US-0227439.
XX
PA   (MEND-) MENDEL BIOTECHNOLOGY INC.
PA   (JIAN/) JIANG C.
PA   (HEAR/) HEARD J.
PA   (PINE/) PINEDA O.
PA   (PILG/) PILGRIM M.
PA   (ADAM/) ADAM L.
PA   (RIEC/) RIECHMANN J L.
PA   (YUGG/) YU G.
PA   (SAMA/) SAMAHA R.
XX
PI   Jiang C,  Heard J,  Pineda O,  Pilgrim M,  Adam L,  Riechmann JL;
PI   Yu G,  Samaha R;
XX
DR   WPI; 2001-335977/35.
DR   P-PSDB; AAE02563.
XX
PT   Nucleic acids encoding plant transcription factor polypeptides, useful
PT   for altering the sugar sensing characteristics of plants and increasing
PT   yield, e.g. corn, potato and cotton plants -
XX
PS   Claim 4; Page 122-124; 151pp; English.
XX
CC   The patent relates to polynucleotides encoding 35 plant transcription
CC   factors which may be used to modify phenotype associated with a plant's
CC   sugar sensing characteristics and increasing yield when their expression
CC   level is altered. Sugars are central regulatory molecules that control
CC   aspects of physiology, metabolism and development. Therefore the cDNAs
CC   and proteins of the invention are useful for modifying the growth and
CC   germination rates of plants, photosynthesis, glyoxylate metabolism,
CC   respiration, starch and sucrose synthesis and degradation, pathogen
CC   response, wounding response, cell cycle regulation, pigmentation,
CC   flowering and senescence of plants and for modifying sink-source
CC   relationships in seeds, tubers, roots, and other storage organs leading
CC   to an increase in yield. The transcription factor polynucleotides and
CC   polypeptides may be used to alter the structure and developmental
CC   characteristics of plants such as soybean, wheat, corn, potato, cotton,
CC   rice, oilseed, sunflower, alfalfa, sugarcane, turf, banana, blackberry,
CC   blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower,
CC   coffee, cucumber, eggplant, grapes, honey dew, lettuce, mango, melon,
CC   onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn,
CC   tobacco, tomato, watermelon, rosaceous fruits and/or vegetable brassicas.
CC   The present sequence is a homolog of Arabidopsis thaliana transcription
CC   factor cDNA.
XX
SQ   Sequence 2240 BP; 685 A; 473 C; 477 G; 605 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 22;  Length 2240;
  Best Local Similarity   80.0%;  Pred. No. 79;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          ||| ||||| |||||||||   |||
Db   1712 GCAATTCGACCCTCGTCAGACACAG 1688


RESULT  16
AAD05746/c
ID   AAD05746 standard; cDNA; 2240 BP.
XX
AC   AAD05746;
XX
DT   31-JUL-2001  (first entry)
XX
DE   Arabidopsis thaliana transcription factor G214 cDNA.
XX
KW   Transcription factor; trait modification; seed characteristic;
KW   structural characteristic; developmental characteristic; gene therapy;
KW   agricultural biotechnology; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             238..2064
FT                   /*tag=  a
FT                   /product= "Transcription factor G214"
XX
PN   WO200135727-A1.
XX
PD   25-MAY-2001.
XX
PF   14-NOV-2000; 2000WO-US31457.
XX
PR   17-NOV-1999;   99US-0166228.
PR   17-APR-2000; 2000US-0197899.
```

```
PR   22-AUG-2000; 2000US-0227439.
XX
PA   (MEND-) MENDEL BIOTECHNOLOGY INC.
PA   (REUB/) REUBER L.
PA   (CREE/) CREELMAN R.
PA   (PILG/) PILGRIM M.
PA   (RIEC/) RIECHMANN J L.
PA   (JIAN/) JIANG C.
PA   (YUGG/) YU G.
PA   (PINE/) PINEDA O.
PA   (HEAR/) HEARD J.
XX
PI   Reuber L,  Creelman R,  Pilgrim M,  Riechmann JL,  Jiang C,  Yu G;
PI   Pineda O,  Heard J;
XX
DR   WPI; 2001-335979/35.
DR   P-PSDB; AAE01911.
XX
PT   Nucleic acids encoding plant transcription factor polypeptides, useful
PT   for altering the developmental and structural characteristics of
PT   plants, e.g. corn, potato and cotton plants -
XX
PS   Claim 4; Page 56-58; 133pp; English.
XX
CC   The present sequence is Arabidopsis thaliana transcription factor G214
CC   cDNA. The transcription factors may be used to modify traits associated
CC   with structural or developmental characteristics of plants, e.g., corn,
CC   potato and cotton plants, when their expression level is altered.
CC   Specifically, they are used for modifying the nutritional content of
CC   plants or seeds, e.g., to modify vitamin and mineral content, to modify
CC   the oil content of seeds, to modify the insoluble sugar content of
CC   seeds, to modify phrenyl lipid content, to modify fatty acid and modify
CC   wax content. They may also be used to alter seed characteristics such
CC   as shelf-life, size, stress tolerance, seedling vigour, pest and
CC   pathogen resistance and germination rate. They may also be used in gene
CC   therapy. Therefore manipulating transcription factor levels in plants
CC   offers great potential in agricultural biotechnology for modifying a
CC   plant's traits.
XX
SQ   Sequence 2240 BP; 685 A; 473 C; 477 G; 605 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 22;  Length 2240;
  Best Local Similarity   80.0%;  Pred. No. 79;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           ||| ||||| |||||||||   |||
Db    1712 GCAATTCGACCCTCGTCAGACACAG 1688


RESULT  17
AAD05773/c
ID   AAD05773 standard; cDNA; 2240 BP.
XX
AC   AAD05773;
XX
DT   31-JUL-2001  (first entry)
XX
DE   Arabidopsis thaliana transcription factor, G214 cDNA.
XX
KW   Transcription factor; biochemical characteristic; controlling element;
KW   structural characteristic; developmental characteristic; gene therapy;
KW   agricultural biotechnology; plant trait modification; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             238..2064
FT                   /*tag=  a
FT                   /product= "Transcription factor, G214"
XX
PN   WO200136597-A1.
XX
PD   25-MAY-2001.
XX
PF   14-NOV-2000; 2000WO-US31344.
XX
PR   17-NOV-1999;   99US-0166228.
PR   17-APR-2000; 2000US-0197899.
PR   22-AUG-2000; 2000US-0227439.
XX
PA   (MEND-) MENDEL BIOTECHNOLOGY INC.
PA   (CREE/) CREELMAN R.
PA   (YUGG/) YU G.
PA   (ADAM/) ADAM L.
PA   (RIEC/) RIECHMANN J L.
PA   (HEAR/) HEARD J.
PA   (SAMA/) SAMAHA R.
PA   (PILG/) PILGRIM M.
PA   (PINE/) PINEDA O.
PA   (JIAN/) JIANG C.
XX
PI   Creelman R,  Yu G,  Adam L,  Riechmann JL,  Heard J,  Samaha R;
PI   Pilgrim M,  Pineda O,  Jiang C;
XX
DR   WPI; 2001-335999/35.
DR   P-PSDB; AAE01889.
XX
PT   Nucleic acids encoding plant transcription factor polypeptides, useful
PT   for altering the biochemical characteristics of plants e.g. corn,
PT   potato and cotton plants -
XX
PS   Claim 4; Page 55-57; 127pp; English.
XX
CC   The present sequence is Arabidopsis thaliana transcription factor,
CC   G214 cDNA. The transcription factor is used for altering a plant's
CC   biochemical characteristics. The transcription factor may be used to
CC   alter the structure and developmental characteristics of plants such as
CC   soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower,
CC   alfalfa, sugar cane, turf, banana, blackberry, blueberry, strawberry,
CC   raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant,
CC   grapes, honey dew, lettuce, mango, melon, onion, papaya, peas, peppers,
CC   pineapple, spinach, squash, sweet corn, tobacco, tomato, watermelon,
CC   rosaceous fruits and/or vegetable brassicas. Transcription factors are
CC   key controlling elements of biological pathways and altering expression
CC   levels of 1 or more transcription factors can change entire biological
CC   pathways in an organism. Therefore manipulating transcription factor
CC   levels in plants offers great potential in agricultural biotechnology
CC   for modifying a plant's traits. Transcription factor cDNA is useful in
CC   gene therapy.
XX
SQ   Sequence 2240 BP; 685 A; 473 C; 477 G; 605 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 22;  Length 2240;
  Best Local Similarity   80.0%;  Pred. No. 79;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           ||| ||||| |||||||||   |||
Db    1712 GCAATTCGACCCTCGTCAGACACAG 1688


RESULT  18
AAV65382/c
ID   AAV65382 standard; cDNA to mRNA; 2254 BP.
XX
AC   AAV65382;
XX
DT   15-FEB-1999  (first entry)
XX
DE   Arabidopsis phytochrome regulated transcription factor CCA1 cDNA.
XX
KW   Phytochrome regulated transcription factor; CCA1;
KW   chlorophyll binding protein; Lhcb1*3; circadian rhythm; flowering;
```

```
KW   transgenic plant; ss.
XX
OS   Arabidopsis thaliana ecotype Columbia.
XX
FH   Key             Location/Qualifiers
FT   CDS             238..2064
FT                   /*tag=  a
XX
PN   WO9848007-A1.
XX
PD   29-OCT-1998.
XX
PF   17-APR-1998;   98WO-US07693.
XX
PR   18-APR-1997;   97US-0843572.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
DR   WPI; 1998-583656/49.
DR   P-PSDB; AAW79280.
XX
PT   Alteration of plant circadian rhythms - by transformation with a
PT   transcription factor for the promoter of the Arabidopsis chlorophyll
PT   binding protein
XX
PS   Claim 1; Page 43-47; 47pp; English.
XX
CC   This cDNA sequence codes for a phytochrome-regulated transcription
CC   factor, designated CCA1 (see AAW79280), that binds to the promoter
CC   region of the chlorophyll binding protein gene (Lhcbl*3) of
CC   Arabidopsis thaliana.  To isolate the clone, a directional cDNA
CC   expression library was constructed in lambda gt22A using cDNA
CC   derived from leaves of Arabidopsis that had been grown in
CC   continuous white light for 3 weeks.  The library was screened with
CC   an A2 fragment of the Lhcbl*3 promoter.  Overlapping phage clones
CC   provided a partial sequence, and the entire 5' end was determined
CC   by primer extension analysis.  The CCA1 cDNA was subsequently used
CC   to identify a genomic clone (see AAV65381).  A claimed method for
CC   altering plant response to daylength comprises transforming a plant
CC   with either the CCA1 genomic DNA or cDNA, a hybridising sequence or
CC   a nucleic acid encoding a protein containing amino acids 24-75 of
CC   the 608-amino acid CCA1 protein.  Manipulation of plant circadian
CC   rhythms (e.g. by overexpression of CCA1) can be used to delay
CC   flowering, and thus increase yield from both forage crops, where
CC   biomass will be produced for longer, and from seed crops such as
CC   rape, where a longer time in vegetative state will lead to larger
CC   plants and thus a higher seed set (claimed).
XX
SQ   Sequence 2254 BP; 699 A; 473 C; 477 G; 605 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 19;  Length 2254;
  Best Local Similarity   80.0%;  Pred. No. 79;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          ||| ||||| |||||||||   |||
Db   1712 GCAATTCGACCCTCGTCAGACACAG 1688


RESULT  19
AAQ15019/c
ID   AAQ15019 standard; DNA; 3234 BP.
XX
AC   AAQ15019;
XX
DT   25-MAR-1992  (first entry)
XX
DE   Encodes mouse bullous pemphigoid antigen protein.
XX
KW   dermatological syndrome; blisters; murine; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2994
FT                   /*tag=  a
XX
PN   JP03206885-A.
XX
PD   10-SEP-1991.
XX
PF   09-APR-1990;   90JP-0093585.
XX
PR   31-OCT-1989;   89JP-0283947.
PR   09-APR-1990;   90JP-0093585.
XX
PA   (TOFU ) TONEN CORP.
XX
DR   WPI; 1991-356566/49.
DR   P-PSDB; AAR15345.
XX
PT   DNA which encodes mouse derived bullous pemphigoid antigen protein -
PT   contg. base sequence which encodes polypeptide composed of specified
PT   996 aminoacid sequence
XX
PS   Claim 1; Fig 1; 23pp; Japanese.
XX
CC   The polypeptide encoded by this nucleotide sequence is recognised
CC   by auto-antibodies in the serum of a bullous pemphigoid patient and
CC   by a human anti-basement membrane zone monoclonal antibody.
CC   See also AAQ15020.
XX
SQ   Sequence 3234 BP; 938 A; 738 C; 873 G; 685 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 12;  Length 3234;
  Best Local Similarity   80.0%;  Pred. No. 83;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          ||  ||||||  || ||||||||||
Db    235 GCTCTTCGATGGTCTTCAGGTGCAG 211


RESULT  20
AAZ30163/c
ID   AAZ30163 standard; DNA; 34094 BP.
XX
AC   AAZ30163;
XX
DT   26-JAN-2000  (first entry)
XX
DE   Complete nucleotide sequence of the PAV-3 genome.
XX
KW   PAV-3; defective recombinant PAV vector; live recombinant virus;
KW   subunit vaccine; nucleic acid immunisation; gene therapy;
KW   genetic disease; hemophilia; cystic fibrosis; cancer; viral infection;
KW   acquired immune deficiency syndrome; PAV antigen; porcine pathogen; ds.
XX
OS   Porcine adenovirus Type 3.
XX
PN   WO9953047-A2.
XX
PD   21-OCT-1999.
XX
PF   15-APR-1999;   99WO-US08220.
XX
PR   15-APR-1998;   98US-0081882.
XX
PA   (UYSA-) UNIV SASKATCHEWAN.
XX
PI   Reddy PS,  Tikoo SK,  Babiuk LA;
XX
DR   WPI; 1999-620422/53.
XX
```

PT New nucleic acids from the genome of porcine adenovirus-3, and derived
PT gene therapy vectors, particularly for immunization -
XX
PS Example 2; Fig 1; 87pp; English.
XX
CC The present sequence represents the complete nucleotide sequence of the
CC genome of porcine adenovirus-3 (PAV-3). The specification also describes
CC a defective recombinant PAV vector comprising inverted terminal repeats
CC (ITR), packaging sequences and at least one heterologous nucleotide
CC sequence (II), but lacking E1 functions. The defective vectors replicate
CC inefficiently in cells (other than helper cells) so are unlikely to be
CC immunogenic. Deletion of the E1 (and optionally other regions) increases
CC the size of heterologous insert that can be packaged. The PAV-3
CC polynucleotides sequences are used to produce (recombinant or defective)
CC vectors that can express heterologous proteins, e.g. for making live,
CC recombinant virus or subunit vaccines, for nucleic acid immunisation or
CC for gene therapy (e.g. of genetic diseases such as hemophilia or cystic
CC fibrosis, cancer, or viral infections, including acquired immune
CC deficiency syndrome), also for in vitro expression of recombinant
CC antigens (for antibody production), antisense RNA, ribozymes or
CC therapeutic proteins. They are also used diagnostically to detect PAV
CC antigens and/or nucleic acid. The vectors may be used in human or
CC veterinary medicine, but particularly for expressing protective
CC determinants of porcine pathogens. Regulatory regions may be used to
CC control expression of heterologous genes. Antibodies raised against PAV-3
CC polypeptides can also be used for diagnosis (to detect PAV-specific
CC antigen).
XX
SQ Sequence 34094 BP; 6240 A; 11070 C; 10693 G; 6091 T; 0 other;

Query Match 66.4%; Score 16.6; DB 20; Length 34094;
Best Local Similarity 82.6%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 atttcgatcctcgtcaggtgcag 25
||||||| | | || |||||||||
Db 27375 ATTTCGTTACGCGGCAGGTGCAG 27353

RESULT 21
AAH67831
ID AAH67831 standard; DNA; 1356 BP.
XX
AC AAH67831;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 2866.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR P-PSDB; AAG92612.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 8; SEQ ID NO: 2866; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 1356 BP; 297 A; 370 C; 350 G; 339 T; 0 other;

Query Match 65.6%; Score 16.4; DB 22; Length 1356;
Best Local Similarity 94.4%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gcatttcgatcctcgtca 18
||||||||||| ||||||
Db 1301 gcatttcgatcatcgtca 1318

RESULT 22
AAH68533/c
ID AAH68533 standard; DNA; 349980 BP.
XX
AC AAH68533;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 7068.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Disclosure; SEQ ID NO: 7068; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein

```
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence 349980 BP; 81250 A; 97718 C; 90621 G; 80391 T; 0 other;


  Query Match             65.6%;  Score 16.4;  DB 22;  Length 349980;
  Best Local Similarity   94.4%;  Pred. No. 2.8e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtca 18
           ||||||||||| ||||||
Db   69211 GCATTTCGATCATCGTCA 69194


RESULT  23
AAC56248/c
ID   AAC56248 standard; DNA; 385 BP.
XX
AC   AAC56248;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Pinus radiata transcription factor DNA sequence #48.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS   Pinus radiata.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M,  McGrath A,  Shenk MA,  Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 1; Page 143; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 385 BP; 128 A; 77 C; 94 G; 86 T; 0 other;


  Query Match             64.8%;  Score 16.2;  DB 21;  Length 385;
  Best Local Similarity   85.7%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 tttcgatcctcgtcaggtgca 24
           |||||||||||| |||| || |
Db      31 TTTCGATCCTCCTCAGCTGGA 11


RESULT  24
AAC57313/c
ID   AAC57313 standard; DNA; 393 BP.
XX
AC   AAC57313;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Pinus radiata transcription factor DNA sequence #674.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS   Pinus radiata.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M,  McGrath A,  Shenk MA,  Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 1; Page 643; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 393 BP; 133 A; 78 C; 95 G; 87 T; 0 other;
```

```
  Query Match             64.8%;  Score 16.2;  DB 21;  Length 393;
  Best Local Similarity   85.7%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 tttcgatcctcgtcaggtgca 24
           ||||||||||| |||| || |
Db      31 TTTCGATCCTCCTCAGCTGGA 11


RESULT  25
AAX30823
ID   AAX30823 standard; DNA; 865 BP.
XX
AC   AAX30823;
XX
DT   20-MAY-1999  (first entry)
XX
DE   Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:100.
XX
KW   Streptococcus pneumoniae strain 0100993; vaccine; immune response;
KW   streptococcal infection; pneumococcal; ss.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9737026-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US05306.
XX
PR   22-AUG-1996;   96US-0025788.
PR   02-APR-1996;   96US-0014690.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Black MT,  Hodgson JE,  Knowles DJC,  Nicholas RO;
PI   Stodola RK;
XX
DR   WPI; 1997-503111/46.
DR   P-PSDB; AAY11240.
XX
PT   Nucleic acids encoding pneumococcal polypeptide(s) - useful in
PT   vaccines, drug screening, etc
XX
PS   Claim 5; Page 119; 354pp; English.
XX
CC   AAX30724 to AAX30946 represent genomic DNA sequences isolated from
CC   Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
CC   encode the novel proteins given in AAY11114 to AAY11367. The proteins,
CC   isolated from Streptococcus pneumoniae, can be used in vaccines against
CC   streptococcal infections and in assays for identifying compounds that
CC   inhibit or activate the activity of the proteins. The antagonists can
CC   be used to treat an individual having need to inhibit a bacterial
CC   protein. Vectors expressing the proteins can be used to induce a
CC   protective immune response in mammals.
XX
SQ   Sequence 865 BP; 270 A; 215 C; 157 G; 223 T; 0 other;


  Query Match             64.8%;  Score 16.2;  DB 18;  Length 865;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atttcgatcctcgtcaggtgc 23
           ||||| | ||| |||||||||
Db     115 atttccagccttgtcaggtgc 135


RESULT  26
AAT98665/c
ID   AAT98665 standard; DNA; 882 BP.
XX
AC   AAT98665;
XX
DT   09-NOV-1998  (first entry)
XX
DE   DNA encoding a S. pneumoniae protein of unknown function.
XX
KW   Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW   immunological response; inoculation; antibody production; inhibitor;
KW   T cell immune response; antimicrobial compound; bacterial adhesion;
KW   extracellular matrix protein; protein-mediated cell invasion; wound;
KW   pathogenesis; ss.
XX
OS   Streptococcus pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   CDS             120..503
FT                   /*tag=  a
XX
PN   WO9743303-A1.
XX
PD   20-NOV-1997.
XX
PF   14-MAY-1997;   97WO-US07950.
XX
PR   14-MAY-1996;   96US-0017670.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Black MT,  Hodgson JE,  Knowles DJC,  Nicholas RO;
PI   Stodola RK;
XX
DR   WPI; 1998-008793/01.
DR   P-PSDB; AAW38617.
XX
PT   Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT   diagnosing anti-microbial agents for treatment of bacterial
PT   infections
XX
PS   Claim 4; Page 182; 483pp; English.
XX
CC   This sequence encodes a Streptococcus pneumoniae protein of unknown
CC   function, and represents a DNA sequence of the invention.
CC   The DNA sequences were isolated from Streptococcus pneumoniae strain
CC   0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC   invention can be used to identify compounds which interact with and
CC   inhibit or activate the activity of the proteins. Antagonists can be
CC   used to treat diseases caused by S. pneumoniae proteins, through genetic
CC   immunisation. They can also be used to induce an immunological response
CC   in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC   of the encoding nucleic acids in a vector adequate to produce antibody
CC   and/or T cell immune responses to protect the animal from disease. The
CC   proteins can also be used to identify antimicrobial compounds which are
CC   capable of inhibiting their bioactivity. In particular the proteins of
CC   the invention can be used to prevent adhesion of bacteria to mammalian
CC   extracellular matrix proteins on in-dwelling devices or in wounds, to
CC   block protein-mediated mammalian cell invasion, and to block the normal
CC   progression of pathogenesis in infections initiated other than by the
CC   implantation of in-dwelling devices or other surgical techniques.
XX
SQ   Sequence 882 BP; 229 A; 162 C; 213 G; 278 T; 0 other;


  Query Match             64.8%;  Score 16.2;  DB 19;  Length 882;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atttcgatcctcgtcaggtgc 23
           ||||| | ||| |||||||||
Db     768 ATTTCCAGCCTTGTCAGGTGC 748
```

```
RESULT  27
AAZ96336
ID   AAZ96336 standard; DNA; 1702 BP.
XX
AC   AAZ96336;
XX
DT   10-APR-2000  (first entry)
XX
DE   S. pneumoniae derived DNA from ORF #164.
XX
KW   Treatment; prevention; disease; diagnosis; gene therapy; screening;
KW   bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9806734-A1.
XX
PD   19-FEB-1998.
XX
PF   15-AUG-1997;   97WO-US14436.
XX
PR   16-AUG-1996;   96US-0024022.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Black MT,  Hodgson JE,  Knowles DJC,  Lonetto MA,  Nicholas RO;
PI   Stodola RK;
XX
DR   WPI; 1998-159452/14.
DR   P-PSDB; AAY85994, AAY85995.
XX
PT   Streptococcus pneumoniae proteins and related DNA - useful for
PT   screening compounds for antibacterial activity
XX
PS   Claim 4; Page 198-199; 640pp; English.
XX
CC   This invention describes novel isolated Streptococcus pneumoniae
CC   polynucleotides (see AAZ96173-Z96494) and their encoded proteins (see
CC   AAY85792-Y86182). The DNA, vectors and host cells described in the
CC   method of the invention are useful for the recombinant expression of the
CC   polypeptides. The polypeptides are useful for treatment or prevention of
CC   disease, or diagnosis of disease related to expression or activity of
CC   such a polypeptide. They can also be used to screen for compounds which
CC   interact with and inhibit or activate such a polypeptide. The
CC   polypeptides (or DNA encoding them, via gene therapy) are also useful
CC   for inducing an immunological response in a mammal. The antagonists are
CC   useful to inhibit such bacterial polypeptides. The polypeptides are
CC   particularly useful to identify antimicrobial compounds and antibiotics.
CC   They are also useful to determine their role in pathogenesis of
CC   infection, dysfunction and disease.
XX
SQ   Sequence 1702 BP; 490 A; 397 C; 348 G; 467 T; 0 other;


  Query Match             64.8%;  Score 16.2;  DB 19;  Length 1702;
  Best Local Similarity   85.7%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atttcgatcctcgtcaggtgc 23
          ||||| | ||| |||||||||
Db    899 atttccagccttgtcaggtgc 919


RESULT  28
AAV37384
ID   AAV37384 standard; DNA; 1702 BP.
XX
AC   AAV37384;
XX
DT   13-OCT-1998  (first entry)
XX
DE   Streptococcus pneumoniae coding region.
XX
KW   coding region; ORF; open reading frame; antibacterial;
KW   infection; prevention; meningitis; ss.
XX
OS   Streptococcus pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1164..1640)
FT                   /*tag=  a
FT                   /product= putative AGAS protein
XX
PN   WO9819689-A1.
XX
PD   14-MAY-1998.
XX
PF   27-OCT-1997;   97WO-US19226.
XX
PR   01-NOV-1996;   96US-0029930.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Black MT,  Hodgson JE,  Knowles DJC,  Lonetto MA,  Nicholas RO;
PI   Reid RH,  Zarfos PN;
XX
DR   WPI; 1998-286586/25.
DR   P-PSDB; AAW60994.
XX
PT   New isolated nucleic acids from Streptococcus pneumoniae - useful,
PT   e.g. for identifying anti-bacterial(s) for treatment and prevention
PT   of meningitis
XX
PS   Claim 1; Page 89-90; 130pp; English.
XX
CC   The sequence is that of a coding region isolated from
CC   S. pneumoniae. Its encoded protein, or agonists of it,
CC   may be useful as an antibacterial for treatment or
CC   prevention of infection, specifically caused by S. pneumoniae
CC   (particularly meningitis) but possibly also Helicobacter
CC   pylori (ulcers and gastric cancer). It may be of particular
CC   use before insertion of an in-dwelling device or any other
CC   invasive procedure. The protein, or nucleic acid encoding
CC   it, can also be used in vaccines to induce a cellular
CC   and/or humoral immune response, or to screen for other
CC   antibacterials. The DNA may also contain flanking sequences
CC   that are potential sources of control elements for bacterial
CC   gene expression. Detecting a sequence encoding the protein
CC   can be used diagnostically, e.g. to detect a mutation for
CC   serotyping or classifying infectious agents.
XX
SQ   Sequence 1702 BP; 490 A; 397 C; 348 G; 467 T; 0 other;


  Query Match             64.8%;  Score 16.2;  DB 19;  Length 1702;
  Best Local Similarity   85.7%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atttcgatcctcgtcaggtgc 23
          ||||| | ||| |||||||||
Db    899 atttccagccttgtcaggtgc 919


RESULT  29
AAV52304/c
ID   AAV52304 standard; DNA; 14736 BP.
XX
AC   AAV52304;
XX
DT   23-OCT-1998  (first entry)
XX
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:171.
XX
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
```

KW computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS Streptococcus pneumoniae.
XX
PN WO9818931-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US19588.
XX
PR 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rosen CA;
XX
DR WPI; 1998-272225/24.
XX
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
XX
PS Claim 1; Page 1085-1094; 1409pp; English.
XX
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 95%
CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridise to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ Sequence 14736 BP; 4665 A; 2568 C; 3255 G; 4247 T; 1 other;

Query Match 64.8%; Score 16.2; DB 19; Length 14736;
Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atttcgatcctcgtcaggtgc 23
|||||.|.|||.|||||||||
Db 6166 ATTTCCAGCCTTGTCAGGTGC 6146

RESULT 30
AAH28893/c
ID AAH28893 standard; DNA; 601 BP.
XX
AC AAH28893;
XX
DT 17-JUL-2001 (first entry)
XX
DE Drosophila melanogaster essential gene fragment, SEQ ID NO: 82.
XX
KW Drosophila melanogaster; fruit fly; essential gene; screening assay;
KW pesticide; crop protection; chromosome 2; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200118547-A1.
XX
PD 15-MAR-2001.
XX
PF 06-SEP-2000; 2000WO-GB03444.
XX
PR 07-SEP-1999; 99GB-0021009.
XX
PA (UNIU) UNIV GLASGOW.
XX
PI Davies RW, Kaiser K, Yang MY;
XX
DR WPI; 2001-281436/29.
XX
PT Screening assays for used for identifying compounds having a
PT physiological effect on proteins identified as being essential -
XX
PS Claim 1; Page 170; 695pp; English.
XX
CC The present sequence is part of an essential gene from Drosophila
CC melanogaster. Lack of expression of the protein encoded by this
CC gene leads to a lethal or semi-lethal phenotype. The invention
CC relates to 902 nucleic acid sequences from genes encoding proteins
CC which are thought to be essential, and to a screening assay for
CC identifying compounds which have a physiological effect on these
CC proteins. Suitable compounds are useful as pesticides and may be used
CC in conjunction with other pesticides and herbicides for crop
CC protection. The gene corresponding to the present sequence is located
CC on chromosome 2.
XX
SQ Sequence 601 BP; 150 A; 159 C; 167 G; 125 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 601;
Best Local Similarity 79.2%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 catttcgatcctcgtcaggtgcag 25
|||||.|||||.|.||.||.||||
Db 473 CATTTGGATCCACTTCTGGAGCAG 450

RESULT 31
AAT59977/c
ID AAT59977 standard; DNA; 1045 BP.
XX
AC AAT59977;
XX
DT 12-MAY-1997 (first entry)
XX
DE 5' untranslated region of the tipE protein coding sequence.
XX
KW Drosophila; tipE; para protein; voltage-dependent cation channel; stroke;
KW pesticide; insecticide; insect; parasitic infection; human; head trauma;
KW neuroprotection; hypoxia; therapy; ss.
XX
OS Drosophila melanogaster.
XX
PN US5593862-A.
XX
PD 14-JAN-1997.
XX
PF 04-OCT-1994; 94US-0317880.
XX
PR 04-OCT-1994; 94US-0317880.
XX
PA (UYNY) UNIV NEW YORK STATE RES FOUND.
XX
PI Feng G, Hall LM;
XX

```
DR   WPI; 1997-099467/09.
XX
PT   Nucleic acid encoding Drosophila melanogaster tipE protein - for
PT   prodn. of recombinant voltage-dependent cation channel
XX
PS   Example 11; Column 35-36; 33pp; English.
XX
CC   AAT59977 and AAT59978 represent the 5' and 3' untranslated regions of
CC   the Drosophila tipE protein (see AAW13843) coding sequence (see
CC   AAT59975). Mutations in the tipE protein, result in a ethyl methane
CC   sulphonate- induced recessive mutation phenotype. Homologous flies for
CC   the mutation paralyse rapidly at 38 degrees, and recover immediately when
CC   returned to 23 degrees. Coexpression of the full length tipE sequence,
CC   and a nucleic acid encoding a para protein results in translation
CC   products that form a functional voltage-dependent cation channel. The
CC   cation channel can be used to screen for pesticides active against
CC   insects such as Drosophila melanogaster and pest insects. The cation
CC   channel can also be used to screen for drugs for use in the treatment and
CC   prevention of parasitic infections in humans and animals, and to screen
CC   drugs for their neuroprotective effect against hypoxia, stroke, and head
CC   trauma.
XX
SQ   Sequence 1045 BP; 379 A; 237 C; 242 G; 187 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 18;  Length 1045;
  Best Local Similarity   79.2%;  Pred. No. 2.2e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           | |||| ||||| | || ||||||
Db     163 CCTTTCTATCCTTGCCACGTGCAG 140


RESULT  32
AAX02016/c
ID   AAX02016 standard; DNA; 1045 BP.
XX
AC   AAX02016;
XX
DT   21-APR-1999  (first entry)
XX
DE   D. melanogaster tipE+ 4Kb clone 5'-UTR.
XX
KW   tipE+; para protein; modulator; voltage dependent cation channel; VDCC;
KW   pesticide; insect control; pharmaceutical agent; neuroprotection;
KW   hypoxia; ischaemia; stroke; head trauma; ss.
XX
OS   Drosophila melanogaster.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..1045
FT                   /*tag=  a
XX
PN   US5871940-A.
XX
PD   16-FEB-1999.
XX
PF   13-JAN-1997;    97US-0782396.
XX
PR   04-OCT-1994;    94US-0317880.
PR   13-JAN-1997;    97US-0782396.
XX
PA   (UYNY ) UNIV NEW YORK STATE RES FOUND.
XX
PI   Feng G,  Hall LM;
XX
DR   WPI; 1999-166632/14.
XX
PT   Screening for agents which modulate ion channel function - using
PT   host cells transformed with nucleic acids encoding the Drosophila
PT   melanogaster tipE and para proteins
XX
PS   Disclosure; Column 35-36; 54pp; English.
XX
CC   This sequence represents the 5'-UTR from a Drosophila melanogaster tipE+
CC   protein which is used in a method for screening for agents which modulate
CC   ion channel function which uses host cells transformed with nucleic acid
CC   encoding Drosophila melanogaster tipE and para proteins. Co-expression of
CC   these genes in the host cell, allows the formation of a functional
CC   voltage dependent cation channel (VDCC) in the cell. The agents
CC   identified can be used as pesticides for the control of Drosophila
CC   melanogaster or other insects. They can also be used to screen
CC   pharmaceutical agents for their neuroprotective affect against e.g.
CC   hypoxia, ischaemia, stroke and head trauma.
XX
SQ   Sequence 1045 BP; 379 A; 237 C; 242 G; 187 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 20;  Length 1045;
  Best Local Similarity   79.2%;  Pred. No. 2.2e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           | |||| ||||| | || ||||||
Db     163 CCTTTCTATCCTTGCCACGTGCAG 140


RESULT  33
AAX36567/c
ID   AAX36567 standard; DNA; 1137 BP.
XX
AC   AAX36567;
XX
DT   07-JUL-1999  (first entry)
XX
DE   Human EDG-3 coding sequence.
XX
KW   EDG-1; EDG-2; EDG-3; EDG-4; EDG-5; PSP-24; human; detection; therapy;
KW   inverse agonist; allosteric modulator; lysophosphatidic acid receptor;
KW   LPA signalling mediated disease; cellular apoptosis; ss.
XX
OS   Homo sapiens.
XX
PN   WO9919513-A2.
XX
PD   22-APR-1999.
XX
PF   09-OCT-1998;    98WO-US21315.
XX
PR   10-OCT-1997;    97US-0061572.
XX
PA   (LXRB-) LXR BIOTECHNOLOGY INC.
XX
PI   Erikson J,  Goddard JG,  Kiefer M;
XX
DR   WPI; 1999-277658/23.
DR   P-PSDB; AAY05490.
XX
PT   Identification of (ant)agonists of LPA receptor EDG-2 for, e.g.
PT   treating LPA signalling mediated diseases such as cellular apoptosis
XX
PS   Claim 1; Page 53-54; 63pp; English.
XX
CC   This sequence encodes human EDG-3.
CC   The invention relates to methods of detecting (ant)agonist, inverse
CC   agonist or allosteric modulators of the lysophosphatidic acid receptors
CC   EDG-1, EDG-2, EDG-3, EDG-4, EDG-5, and PSP-24. The methods are used to
CC   identify (ant)agonists and allosteric modulator of the lysophosphatidic
CC   acid (LPA) EDG2 receptor, e.g. to treat LPA signalling mediated disease
CC   such cellular apoptosis.
XX
SQ   Sequence 1137 BP; 226 A; 362 C; 298 G; 251 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 20;  Length 1137;
```

```
  Best Local Similarity   79.2%;  Pred. No. 2.2e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           ||||| |||| | |||| |||| |
Db     429 CATTTTGATCATTGTCAAGTGCCG 406


RESULT  34
AAZ22300/c
ID   AAZ22300 standard; cDNA; 1137 BP.
XX
AC   AAZ22300;
XX
DT   28-MAR-2000  (first entry)
XX
DE   Human EDG3sb polypeptide encoding cDNA.
XX
KW   EDG3sb protein; infection; HIV-1; HIV-2; pain; cancer; diabetes; obesity;
KW   anorexia; bulimia; asthma; Parkinson's disease; acute heart failure;
KW   hypotension; hypertension; urinary retention; osteoporosis; allergy;
KW   angina pectoris; myocardial infarction; stroke; ulcer; psychotic; human;
KW   neurological disorder; dyskinesia; Huntington's disease; antidepressant;
KW   nootropic; antimigraine; antiemetic; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1137
FT                   /*tag= a
FT                   /transl_except= (pos:568..570, aa: Pro)
FT                   /product= "EDG3sb polypeptide"
XX
PN   WO9960019-A1.
XX
PD   25-NOV-1999.
XX
PF   12-MAY-1999;   99WO-US10337.
XX
PR   20-MAY-1998;   98US-0082088.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Tsui P;
XX
DR   WPI; 2000-116310/10.
DR   P-PSDB; AAY59176.
XX
PT   Novel polypeptides, used to develop products for treating e.g.
PT   infections, cancers, diabetes, obesity, asthma, Parkinson's disease,
PT   heart failure, stroke or neurological disorders  -
XX
PS   Claim 13; Page 27; 36pp; English.
XX
CC   This cDNA encodes a human EDG3sb polypeptide. The EDG3sb protein can be
CC   expressed by standard recombinant methodology. The products can be used
CC   for treating e.g. infections such as bacterial, fungal, protozoan and
CC   viral infections, particularly infections caused by HIV-1 or HIV-2, pain,
CC   cancers, diabetes, obesity, anorexia, bulimia, asthma, Parkinson's
CC   disease, acute heart failure, hypotension, hypertension, urinary
CC   retention, osteoporosis, angina pectoris, myocardial infarction, stroke,
CC   ulcers, allergies, benign prostatic hypertrophy, migraine, vomiting,
CC   psychotic and neurological disorders, including anxiety, schizophrenia,
CC   manic depression, depression, delirium, dementia, and severe mental
CC   retardation, and dyskinesias, such as Huntington's disease or Gilles dela
CC   Tourett's syndrome, related to either an excess of, or an under-
CC   expression of EGD3sb polypeptide activity. The products can also be used
CC   for detection, diagnosis and drug screening.
XX
SQ   Sequence 1137 BP; 225 A; 362 C; 299 G; 251 T; 0 other;

  Query Match           64.0%;  Score 16;  DB 21;  Length 1137;
```

```
  Best Local Similarity   79.2%;  Pred. No. 2.2e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           ||||| |||| | |||| |||| |
Db     429 CATTTTGATCATTGTCAAGTGCCG 406


RESULT  35
AAS00258/c
ID   AAS00258 standard; DNA; 1137 BP.
XX
AC   AAS00258;
XX
DT   11-MAY-2001  (first entry)
XX
DE   LPA receptor-related DNA sequence #2.
XX
KW   LPA receptor; EDG-2; lysophosphatidic acid; phospholipid; tumour;
KW   cell signalling; MAP kinase; LPA modulator; neurodegenerative disease;
KW   Alzheimer's disease; Parkinson's disease; neuron damage; apoptosis;
KW   ischaemic heart disease; viral; HIV; inflammatory bowel disease;
KW   organ transplant; human; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1137
FT                   /*tag= a
XX
PN   WO200112838-A2.
XX
PD   22-FEB-2001.
XX
PF   11-AUG-2000; 2000WO-US22101.
XX
PR   18-AUG-1999;   99US-0376399.
XX
PA   (ATAI-) ATAIRGIN TECHNOLOGIES INC.
XX
PI   Erickson J,  Goddard JG,  Kiefer M,  Picker D;
XX
DR   WPI; 2001-226550/23.
XX
PT   Modulating activity of a lysophosphatidic acid or its receptor for
PT   treating tumours, viral diseases, involves introducing to LPA or its
PT   receptor a composition containing diol, epoxide or phosphate compound
PT   -
XX
PS   Disclosure; Page 78-79; 89pp; English.
XX
CC   The sequence represents the LPA receptor-related sequence #2 used during
CC   analysis of lysophosphatidic acid receptor (LPA), EDG-2. The sequence is
CC   given in the specification but no further information is given. LPA is a
CC   phospholipid found in a variety of plant and animal products. EDG-2 is
CC   involved in cell signalling through activation of a MAP kinase cascade-
CC   dependent reporter. Modulating the activity of a lysophosphatidic acid
CC   (LPA) or LPA receptor (EDG-2) involves introducing a composition
CC   comprising LPA modulators to the LPA or receptor. The method is useful
CC   for treating diseases characterised by slowed growth or repair of
CC   neuronal cells, neurodegenerative diseases, such as Alzheimer's disease,
CC   Parkinson's disease, and acute neuron damage, for modulating apoptotic
CC   pathways and treating ischaemic heart disease, tumours, viral diseases
CC   bowel disease, and rejection of organ transplants.
XX
SQ   Sequence 1137 BP; 226 A; 362 C; 298 G; 251 T; 0 other;

  Query Match           64.0%;  Score 16;  DB 22;  Length 1137;
  Best Local Similarity   79.2%;  Pred. No. 2.2e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
```

```
          ||||| |||| | |||| |||| |
Db    429 CATTTTGATCATTGTCAAGTGCCG 406


RESULT  36
AAF84497/c
ID   AAF84497 standard; cDNA; 1140 BP.
XX
AC   AAF84497;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Rat G protein-coupled receptor EDG3-encoding cDNA.
XX
KW   Rat; EDG3; G protein-coupled receptor; GPCR; 7TM receptor;
KW   7 transmembrane domain receptor; infection; viral; bacterial; fungal;
KW   protozoan; HIV-1; HIV-2; pain; cancer; diabetes; obesity; anorexia;
KW   bulimia; osteoporosis; asthma; allergy; urinary retention;
KW   acute heart failure; hypotension; hypertension; angina pectoris;
KW   myocardial infarction; stroke; ulcer; migraine; vomiting;
KW   psychotic disorder; neurological disorder; anxiety; schizophrenia;
KW   manic depression; bipolar disorder; depression; delirium; dementia;
KW   severe mental retardation; dyskinesia; Parkinson's disease;
KW   Huntington's disease; Gilles de la Tourette's syndrome;
KW   transgenic animal; drug screening; signal transduction;
KW   agonist; antagonist; ss.
XX
OS   Rattus norvegicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1140
FT                   /*tag=  a
FT                   /product= "Rat EDG3"
XX
PN   WO200127137-A2.
XX
PD   19-APR-2001.
XX
PF   12-OCT-2000; 2000WO-US28231.
XX
PR   12-OCT-1999;   99US-0159019.
PR   11-OCT-2000; 2000US-0686105.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Lane P,  Tsui P,  Elshourbagy N;
XX
DR   WPI; 2001-281971/29.
DR   P-PSDB; AAB73483.
XX
PT   Rattus norvegicus EDG3 polynucleotide, useful for screening antagonists
PT   and agonists which can be used in the treatment of infections (e.g.
PT   bacterial, fungal, protozoan and viral infections), pain, cancers,
PT   diabetes, obesity and anorexia -
XX
PS   Claim 1; Page 30; 31pp; English.
XX
CC   This sequence represents cDNA encoding the rat EDG3 protein. The rat EDG3
CC   protein is a member of the G protein-coupled receptor (GPCR) superfamily,
CC   and like all GPCRs, EDG3 has 7 putative transmembrane domains and is
CC   involved in signal transduction. The invention also relates to expression
CC   vectors and host cells comprising EDG3 DNA, to recombinant expression of
CC   EDG3, drug screening methods that use rat EDG3, and to EDG3 agonists and
CC   antagonists. Rat EDG3 nucleic acid sequences may be used to isolate cDNA
CC   and/or genomic clones encoding EDG3 homologues or orthologues,
CC   particularly from other species, and may also be used in chromosome
CC   localisation studies. EDG3 nucleotides may be used to generate transgenic
CC   animals, including knockout animals, in which the function of EDG3 can be
CC   elucidated. Such animal models may provide an insight into treating a
CC   variety of human disorders. Such disorders include bacterial, fungal,
CC   protozoal and viral infections, particularly HIV-1 or HIV-2 infections;
CC   pain; cancers; benign prostatic hypertrophy; diabetes; obesity; anorexia;
CC   bulimia; osteoporosis; asthma; allergies; urinary retention; acute heart
CC   failure; hypotension; hypertension; angina pectoris; myocardial
CC   infarction; stroke; ulcers; migraine; vomiting; psychotic and
CC   neurological disorders such as anxiety, schizophrenia, manic depression,
CC   depression, delirium, dementia, and severe mental retardation, and
CC   dyskinesias, such as Parkinson's disease, Huntington's disease or Gilles
CC   de la Tourette's syndrome. EDG3 proteins and nucleotides may also be used
CC   to screen compounds for their ability to modulate EDG3 activity or
CC   expression.
XX
SQ   Sequence 1140 BP; 234 A; 343 C; 289 G; 274 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 22;  Length 1140;
  Best Local Similarity   79.2%;  Pred. No. 2.2e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           ||| | |||| | |||||||||| |
Db     432 CATCTTGATCATGGTCAGGTGCCG 409


RESULT  37
AAA65913/c
ID   AAA65913 standard; DNA; 1152 BP.
XX
AC   AAA65913;
XX
DT   05-OCT-2000  (first entry)
XX
DE   E. coli proliferation associated coding sequence SEQ ID NO:105.
XX
KW   Escherichia coli; E. coli; proliferation; inhibition; screening;
KW   antimicrobial; bacterial growth; antisense therapy; antibacterial; ds.
XX
OS   Escherichia coli.
XX
PN   WO200044906-A2.
XX
PD   03-AUG-2000.
XX
PF   27-JAN-2000; 2000WO-US02200.
XX
PR   27-JAN-1999;   99US-0117405.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Zyskind J,  Ohlsen KL,  Trawick J,  Forsyth RA,  Froelich JM,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2000-514822/46.
DR   P-PSDB; AAB15908.
XX
PT   Novel polynucleotides and polypeptides associated with microorganism
PT   proliferation, used to identify inhibitors of bacterial growth and
PT   proliferation, for use in antisense therapy -
XX
PS   Claim 8; Page 123; 316pp; English.
XX
CC   AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
CC   sequences derived from Escherichia coli which inhibit E. coli
CC   proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent
CC   nucleotide and protein sequences associated with E. coli proliferation.
CC   AAA66056 and AAA66057 represent primers used for sequencing E. coli
CC   proliferation inhibiting nucleotide inserts in an example from the
CC   present invention. Methods from the present invention can be used to
CC   identify a proliferation- required gene in a microorganism, by contacting
CC   a microorganism with a proliferation-required gene activity inhibitory
CC   nucleic acid identified in another organism, and determining if
CC   inhibition occurs in the second microorganism. The nucleic acid sequences
CC   identified as being required for bacterial growth and proliferation, can
CC   be used for antisense therapy for killing bacteria.
XX
```

```
SQ   Sequence 1152 BP; 267 A; 270 C; 354 G; 261 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 21;  Length 1152;
  Best Local Similarity   79.2%;  Pred. No. 2.2e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgca 24
           ||||  |||||||| ||||  |||
Db     260 GCATCGCGATCCTCATCAGCCGCA 237


RESULT  38
AAA39397/c
ID   AAA39397 standard; DNA; 1152 BP.
XX
AC   AAA39397;
XX
DT   15-SEP-2000  (first entry)
XX
DE   Escherichia coli monomeric FtsZ nucleotide sequence SEQ ID NO:5.
XX
KW   Staphylococcus aureus; monomeric FtsZ; monomeric FtsA; antibacterial;
KW   Escherichia coli; multimeric protein; bacterial infection; vaccine;
KW   microbial infection; screening; ds.
XX
OS   Escherichia coli.
XX
PN   WO200024252-A1.
XX
PD   04-MAY-2000.
XX
PF   20-OCT-1999;   99WO-US24653.
XX
PR   23-OCT-1998;   98US-0105315.
PR   19-APR-1999;   99US-0129965.
PR   03-MAY-1999;   99US-0132333.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Yan K,  Pearce KH;
XX
DR   WPI; 2000-350517/30.
DR   P-PSDB; AAY91110.
XX
PT   Multimeric FtsZ:FtsA polypeptide, useful as a vaccine for treating
PT   bacterial infections and for diagnosing diseases associated with
PT   bacterial infection  -
XX
PS   Disclosure; Page 4; 45pp; English.
XX
CC   The present invention describes isolated multimeric FtsZ:FtsA
CC   polypeptides (I) comprising the protein sequences given in AAY91108,
CC   AAY91109, AAY91110 and AAY91111. Also described are: (1) an antibody
CC   (Ab) specific for (I); (2) a method (M1) for treating an individual
CC   requiring enhanced or reduced activity or expression of (I) comprising
CC   administering an agonist or antagonist to (I), respectively; (3) a
CC   method (M2) for diagnosing a disease or susceptibility to a disease
CC   related to expression or activity of (I) comprising determining the
CC   presence or absence of a mutation in the nucleotide sequence encoding
CC   (I) and/or analysing for the presence or quantity of (I) in a sample;
CC   and (4) a method (M3) for screening to identify compounds that activate
CC   or inhibit the function of (I) selected from: (a) measuring the binding
CC   of a candidate compound to the polypeptide using a label; (b) measuring
CC   the binding of a candidate compound to (I) in the presence of a labeled
CC   competitor; (c) testing whether the candidate compound results in a
CC   signal generated by activation or inhibition of (I); or (d) detecting
CC   the effect of a candidate compound on the production of mRNA encoding
CC   (I) using e.g. enzyme linked immunosorbent assay (ELISA). AAY91108 to
CC   AAY91111 are useful as vaccines for treating microbial infections such
CC   as Staphylococcus, Streptococcus and Escherichia coli infections. The
CC   polypeptides are also useful for screening for antibacterial compounds
CC   and for detecting diseases associated with microbial infections. The
CC   present sequence encodes Escherichia coli FtsZ.
XX
SQ   Sequence 1152 BP; 267 A; 270 C; 354 G; 261 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 21;  Length 1152;
  Best Local Similarity   79.2%;  Pred. No. 2.2e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgca 24
           ||||  |||||||| ||||  |||
Db     260 GCATCGCGATCCTCATCAGCCGCA 237


RESULT  39
AAX59380/c
ID   AAX59380 standard; DNA; 1792 BP.
XX
AC   AAX59380;
XX
DT   20-SEP-1999  (first entry)
XX
DE   Aspergillus nidulans essential gene AN85.
XX
KW   AN85; essential gene; fungicide; fungistatic; screening; infection;
KW   aspergillosis; therapy; ds.
XX
OS   Aspergillus nidulans.
XX
FH   Key             Location/Qualifiers
FT   CDS             230..1478
FT                   /*tag=  a
FT                   /note= "contains introns"
FT   exon            230..309
FT                   /*tag=  b
FT                   /number= 1
FT   intron          310..374
FT                   /*tag=  c
FT                   /number= 1
FT   exon            375..818
FT                   /*tag=  d
FT                   /number= 2
FT   intron          819..875
FT                   /*tag=  e
FT                   /number= 2
FT   exon            876..1149
FT                   /*tag=  f
FT                   /number= 3
FT   intron          1150..1199
FT                   /*tag=  g
FT                   /number= 3
FT   exon            1200..1478
FT                   /*tag=  h
FT                   /number= 4
XX
PN   WO9924580-A2.
XX
PD   20-MAY-1999.
XX
PF   09-NOV-1998;   98WO-US23874.
XX
PR   07-NOV-1997;   97US-0965762.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Gavrias V,  Koltin Y;
XX
DR   WPI; 1999-418430/35.
DR   P-PSDB; AAY06417.
XX
PT   Aspergillus nidulans nucleic acids encoding essential proteins AN97,
PT   AN80, AN85 and AN17
XX
```

PS Claim 2i; Fig 3; 78pp; English.
XX
CC This is the nucleotide sequence of Aspergillus nidulans essential
CC gene AN85 (ATCC 209473). AN85 is one of four new genes (see
CC AAX59378-81) discovered in A. aspergillus that are essential for
CC survival. The 4 genes were identified by preparing a library of
CC approximately 1,000 A. nidulans mutants using 4-nitroquinoline as
CC mutagen, and screening the mutants for conditional sensitivity to
CC temperature. Genes encoding AN97, AN80, AN85 and AN17 are useful
CC molecular tools for identifying similar genes in pathogenic
CC microorgansms. AN polypeptides and the essential genes encoding
CC them are useful targets for identifying antifungal agents that are
CC inhibitors of the pathogens in which the AN polypeptides are
CC expressed. These antifungal agents are useful for treating fungal
CC infections such as aspergillosis. Claimed methods for treating
CC a fungal infection involve administering a pharmaceutical
CC composition comprising an antifungal agent, or a ribozyme or
CC antisense nucleic acid that inhibits the function of an AN
CC polypeptide.
XX
SQ Sequence 1792 BP; 447 A; 479 C; 435 G; 431 T; 0 other;

Query Match 64.0%; Score 16; DB 20; Length 1792;
Best Local Similarity 79.2%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 catttcgatcctcgtcaggtgcag 25
||||||||| | | |||||||||
Db 708 CATTTCGAACATGCGCAGGTGCAG 685

RESULT 40
AAI59340
ID AAI59340 standard; cDNA; 2489 BP.
XX
AC AAI59340;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1543.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM40184.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 1543; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2489 BP; 722 A; 604 C; 550 G; 613 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 2489;
Best Local Similarity 79.2%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 catttcgatcctcgtcaggtgcag 25
||| || |||| ||||||||||
Db 312 catgtctatcccgctcaggtgcag 335

RESULT 41
AAZ52889
ID AAZ52889 standard; cDNA; 2567 BP.
XX
AC AAZ52889;
XX
DT 14-MAR-2000 (first entry)
XX
DE Human prostate tumor cDNA library derived EST fragment #32.
XX
KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW treatment; ds.
XX
OS Homo sapiens.
XX
PN DE19820190-A1.
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
PR 28-APR-1998; 98DE-1020190.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR WPI; 1999-621386/54.
DR P-PSDB; AAY73907, AAY73908, AAY73909.
XX
PT New human nucleic acid sequences from pancreatic tumors, and related
PT proteins -
XX
PS Claim 2; Page 207-208; 502pp; German.
XX
CC This invention describes novel polypeptides and their encoding nucleic
CC acids derived from human pancreatic tumor tissue which have cytostatic

```
CC   activity. The sequences are also useful in producing pharmaceutical
CC   compositions for treatment of pancreatic tumors. AAZ52858-Z53014
CC   represent expressed sequence tag (EST) fragments derived from a human
CC   pancreatic tumor cDNA library and which encode the proteins represented
CC   in AAY73814-Y74252.
XX
SQ   Sequence 2567 BP; 459 A; 833 C; 731 G; 544 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 20;  Length 2567;
  Best Local Similarity   79.2%;  Pred. No. 2.4e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           |||||| ||||  | ||||| |||
Db     371 catttccatccatgacaggtacag 394


RESULT  42
AAQ23963/c
ID   AAQ23963 standard; DNA; 3708 BP.
XX
AC   AAQ23963;
XX
DT   26-OCT-1992  (first entry)
XX
DE   Probe BtY2.
XX
KW   BtY1; BtY2; nucleic acid hybridisation; sex determination; ss.
XX
OS   Bos taurus.
XX
PN   WO9206215-A.
XX
PD   16-APR-1992.
XX
PF   03-OCT-1991;   91WO-CA00353.
XX
PR   04-OCT-1990;   90CA-2026926.
XX
PA   (UYCA-) UNIV CALGARY.
PA   (WINK/) WINKFEIN R J.
XX
PI   Dixon GH,  Winkfein RJ;
XX
DR   WPI; 1992-150898/18.
XX
PT   Y-chromosome specific nucleic acid sequences - used for
PT   determining the sex of bovine or ruminant embryo, fetus or sex
PT   chromosome content of sperm or other cells
XX
PS   Disclosure; Page 34; 51pp; English.
XX
CC   The sequences given in AAQ23962 and AAQ23963 are both sensitive Y-
CC   chromosome probes, designated BtY1 and BtY2 respectively.  BtY1 is a
CC   1.859 kb PstI fragment whereas BtY2 is a 3.71 kb SacI fragment, both
CC   of which were cloned from bovine DNA.  Using these probes it is
CC   possible to sex ruminant/bovine embryos by nucleic acid hybridisation,
CC   rapidly and with virtually 100% accuracy.  The amount of DNA required
CC   is equal to the amount obtd. from two or fewer embryonic cells.  The
CC   probes can be used for determining the sex of an embryo, foetus or
CC   the sex chromosome content of sperm or other cells.  The probes can
CC   also be used to detect variations in amts. and/or variations in
CC   sequence of corresponding sequences in individual mammals, eg. in
CC   paternity testing of male offspring.
XX
SQ   Sequence 3708 BP; 883 A; 1053 C; 798 G; 973 T; 1 other;


  Query Match             64.0%;  Score 16;  DB 13;  Length 3708;
  Best Local Similarity   79.2%;  Pred. No. 2.6e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgca 24
           |||  ||| |||| |||||| |||
Db    1533 GCAGATCGGTCCTTGTCAGGAGCA 1510


RESULT  43
AAA53179/c
ID   AAA53179 standard; DNA; 3904 BP.
XX
AC   AAA53179;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Streptococcus bovis plasmid pSB01 nucleotide sequence SEQ ID NO:1.
XX
KW   Streptococcus bovis; autonomously replicating sequence; ARS;
KW   colibacillus; shuttle vector; lumen microorganism; ds.
XX
OS   Streptococcus bovis.
XX
PN   JP3044303-B1.
XX
PD   22-MAY-2000.
XX
PF   07-DEC-1998;   98JP-0347017.
XX
PR   07-DEC-1998;   98JP-0347017.
XX
PA   (NORQ ) NORINSUISANSHO CHIKUSAN SHIKENJOCHO.
XX
DR   WPI; 2000-415207/36.
XX
PT   Plasmid used as a shuttle vector between Streptococcus bovis and
PT   Colibacillus contains autonomously replicating sequence with specified
PT   base pairs -
XX
PS   Claim 2; Page 5-6; 9pp; Japanese.
XX
CC   The present invention describes a plasmid used as a shuttle vector
CC   between Streptococcus bovis and Colibacillus, which contains an
CC   autonomously replicating sequence (ARS) of S. bovis origin with a
CC   polynucleotide sequence of 3904 base pairs (the present sequence), or
CC   the DNA which is hybridised under stringent conditions to the
CC   polynucleotide sequence. The plasmid acts as a shuttle vector between
CC   Streptococcus bovis and Colibacillus.
XX
SQ   Sequence 3904 BP; 1447 A; 554 C; 743 G; 1160 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 21;  Length 3904;
  Best Local Similarity   79.2%;  Pred. No. 2.6e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgca 24
           || |||  ||| |||| |||||||
Db    2721 GCTTTTAAATCTTCGTTAGGTGCA 2698


RESULT  44
AAT59975/c
ID   AAT59975 standard; DNA; 3954 BP.
XX
AC   AAT59975;
XX
DT   12-MAY-1997  (first entry)
XX
DE   Full length tipE protein coding sequence.
XX
KW   Drosophila; tipE; para protein; voltage-dependent cation channel; stroke;
KW   pesticide; insecticide; insect; parasitic infection; human; head trauma;
KW   neuroprotection; hypoxia; therapy; ss.
XX
OS   Drosophila melanogaster.
```

```
XX
FH   Key             Location/Qualifiers
FT   CDS             1046..2404
FT                   /*tag=  a
FT                   /product= tipE
XX
PN   US5593862-A.
XX
PD   14-JAN-1997.
XX
PF   04-OCT-1994;   94US-0317880.
XX
PR   04-OCT-1994;   94US-0317880.
XX
PA   (UYNY ) UNIV NEW YORK STATE RES FOUND.
XX
PI   Feng G,  Hall LM;
XX
DR   WPI; 1997-099467/09.
DR   P-PSDB; AAW13843.
XX
PT   Nucleic acid encoding Drosophila melanogaster tipE protein - for
PT   prodn. of recombinant voltage-dependent cation channel
XX
PS   Disclosure; Column 29-32; 33pp; English.
XX
CC   This sequence represents the full length coding sequence for the
CC   Drosophila tipE protein.  Mutations in the tipE protein, result in a
CC   ethyl methane sulphonate-induced recessive mutation phenotype.
CC   Homologous flies for the mutation paralyse rapidly at 38 degrees, and
CC   recover immediately when returned to 23 degrees.  Coexpression of this
CC   sequence, and a nucleic acid encoding a para protein results in
CC   translation products that form a functional voltage-dependent cation
CC   channel.  The cation channel can be used to screen for pesticides active
CC   against insects such as Drosophila melanogaster and pest insects.  The
CC   cation channel can also be used to screen for drugs for use in the
CC   treatment and prevention of parasitic infections in humans and animals,
CC   and to screen drugs for their neuroprotective effect against hypoxia,
CC   stroke, and head trauma.
XX
SQ   Sequence 3954 BP; 1293 A; 960 C; 867 G; 834 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 18;  Length 3954;
  Best Local Similarity   79.2%;  Pred. No. 2.6e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           | |||| ||||| | || ||||||
Db     163 CCTTTCTATCCTTGCCACGTGCAG 140


RESULT  45
AAX02015/c
ID   AAX02015 standard; DNA; 3954 BP.
XX
AC   AAX02015;
XX
DT   21-APR-1999  (first entry)
XX
DE   D. melanogaster tipE+ genomic DNA.
XX
KW   tipE+; para protein; modulator; voltage dependent cation channel; VDCC;
KW   pesticide; insect control; pharmaceutical agent; neuroprotection;
KW   hypoxia; ischaemia; stroke; head trauma; ss.
XX
OS   Drosophila melanogaster.
XX
FH   Key             Location/Qualifiers
FT   CDS             1046..2404
FT                   /*tag=  a
FT                   /product= "tipE+"
XX
PN   US5871940-A.
XX
PD   16-FEB-1999.
XX
PF   13-JAN-1997;   97US-0782396.
XX
PR   04-OCT-1994;   94US-0317880.
PR   13-JAN-1997;   97US-0782396.
XX
PA   (UYNY ) UNIV NEW YORK STATE RES FOUND.
XX
PI   Feng G,  Hall LM;
XX
DR   WPI; 1999-166632/14.
DR   P-PSDB; AAW92459.
XX
PT   Screening for agents which modulate ion channel function - using
PT   host cells transformed with nucleic acids encoding the Drosophila
PT   melanogaster tipE and para proteins
XX
PS   Disclosure; Fig 7A-B; 54pp; English.
XX
CC   This sequence encodes a Drosophila melanogaster tipE+ protein which is
CC   used in a method for screening for agents which modulate ion channel
CC   function which uses host cells transformed with nucleic acid encoding
CC   Drosophila melanogaster tipE and para proteins. Co-expression of these
CC   genes in the host cell, allows the formation of a functional voltage
CC   dependent cation channel (VDCC) in the cell. The agents identified
CC   can be used as pesticides for the control of Drosophila melanogaster or
CC   other insects. They can also be used to screen pharmaceutical agents for
CC   their neuroprotective affect against e.g. hypoxia, ischaemia, stroke and
CC   head trauma.
XX
SQ   Sequence 3954 BP; 1293 A; 960 C; 867 G; 834 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 20;  Length 3954;
  Best Local Similarity   79.2%;  Pred. No. 2.6e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           | |||| ||||| | || ||||||
Db     163 CCTTTCTATCCTTGCCACGTGCAG 140


Search completed: December 26, 2001, 12:19:34
Job time: 6935 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:27:04 ; Search time 141.92 Seconds
(without alignments)
39.895 Million cell updates/sec

Title: US-09-396-196F-8
Perfect score: 25
Sequence: 1 gcatttcgatcctcgtcaggtgcag 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
	2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
	3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
	4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
	5	17.2	68.8	876	1	US-08-726-725-5	Sequence 5, Appli
c	6	16	64.0	1045	1	US-08-317-880-3	Sequence 3, Appli
c	7	16	64.0	1045	2	US-08-782-396-3	Sequence 3, Appli
c	8	16	64.0	1137	3	US-09-082-088-1	Sequence 1, Appli
c	9	16	64.0	1792	4	US-08-965-762-7	Sequence 7, Appli
c	10	16	64.0	3713	1	US-08-330-537-3	Sequence 3, Appli
c	11	16	64.0	3713	1	US-08-330-537-4	Sequence 4, Appli
c	12	16	64.0	3954	1	US-08-317-880-1	Sequence 1, Appli
c	13	16	64.0	3954	2	US-08-782-396-1	Sequence 1, Appli
c	14	16	64.0	4928	1	US-08-399-561-1	Sequence 1, Appli
c	15	15.6	62.4	996	4	US-08-875-533-59	Sequence 59, Appl
c	16	15.6	62.4	1139	2	US-08-933-750C-71	Sequence 71, Appl
c	17	15.6	62.4	1139	3	US-09-234-613-71	Sequence 71, Appl
	18	15.4	61.6	516	1	US-07-841-646-8	Sequence 8, Appli
	19	15.4	61.6	516	1	US-08-147-023-8	Sequence 8, Appli
	20	15.4	61.6	516	1	US-08-447-570-8	Sequence 8, Appli
	21	15.4	61.6	516	2	US-08-449-700-8	Sequence 8, Appli
	22	15.4	61.6	516	2	US-08-449-699A-8	Sequence 8, Appli
	23	15.4	61.6	1407	4	US-09-193-377B-2	Sequence 2, Appli
	24	15.4	61.6	1420	4	US-09-193-377B-4	Sequence 4, Appli
	25	15.4	61.6	1428	4	US-09-193-377B-1	Sequence 1, Appli
c	26	15.4	61.6	3933	1	US-08-199-776-1	Sequence 1, Appli
c	27	15.4	61.6	3933	3	US-08-663-731-1	Sequence 1, Appli
c	28	15.4	61.6	3933	3	US-08-879-338-1	Sequence 1, Appli
c	29	15.4	61.6	3933	5	PCT-US95-02044-1	Sequence 1, Appli
	30	15.4	61.6	8931	3	US-09-051-019-1	Sequence 1, Appli
c	31	15.2	60.8	340	3	US-08-441-971-16	Sequence 16, Appl
c	32	15.2	60.8	340	4	US-08-221-653-16	Sequence 16, Appl
c	33	15.2	60.8	340	4	US-08-442-144A-16	Sequence 16, Appl
c	34	15.2	60.8	340	4	US-08-441-970-16	Sequence 16, Appl
c	35	15.2	60.8	1704	1	US-08-528-199-2	Sequence 2, Appli
c	36	15.2	60.8	1704	1	US-08-528-199-5	Sequence 5, Appli
c	37	15	60.0	27	1	US-08-758-306-140	Sequence 140, App
c	38	15	60.0	27	1	US-08-758-306-716	Sequence 716, App
c	39	15	60.0	27	1	US-08-758-306-1112	Sequence 1112, Ap
	40	15	60.0	233	4	US-09-060-756-714	Sequence 714, App
c	41	15	60.0	666	4	US-08-998-416-995	Sequence 995, App
c	42	15	60.0	707	4	US-08-998-416-1093	Sequence 1093, Ap
c	43	15	60.0	2330	2	US-08-557-128-3	Sequence 3, Appli
c	44	15	60.0	2330	4	US-09-242-690A-34	Sequence 34, Appl
	45	15	60.0	2729	1	US-07-938-782A-1	Sequence 1, Appli
	46	15	60.0	2729	1	US-08-630-524-1	Sequence 1, Appli
	47	15	60.0	2729	5	PCT-US93-08131-1	Sequence 1, Appli
c	48	15	60.0	3505	1	US-08-454-097-34	Sequence 34, Appl
c	49	15	60.0	3505	1	US-08-468-036-43	Sequence 43, Appl
c	50	15	60.0	3505	2	US-08-376-843-43	Sequence 43, Appl
c	51	15	60.0	3505	3	US-08-185-359-34	Sequence 34, Appl
	52	15	60.0	15378	3	US-08-785-420-1	Sequence 1, Appli
	53	15	60.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	54	15	60.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	55	15	60.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	56	14.8	59.2	61	3	US-09-023-228B-59	Sequence 59, Appl
	57	14.8	59.2	1362	2	US-08-851-088-7	Sequence 7, Appli
	58	14.8	59.2	1362	3	US-08-851-089-1	Sequence 1, Appli
	59	14.8	59.2	4144	3	US-08-851-089-12	Sequence 12, Appl
c	60	14.8	59.2	4144	3	US-08-851-089-13	Sequence 13, Appl
	61	14.8	59.2	5036	4	US-09-177-349-2	Sequence 2, Appli
c	62	14.6	58.4	27	3	US-08-985-162-1006	Sequence 1006, Ap
c	63	14.6	58.4	29	3	US-08-574-396-33	Sequence 33, Appl
c	64	14.6	58.4	29	3	US-08-488-181-31	Sequence 31, Appl
c	65	14.6	58.4	29	4	US-08-973-568-33	Sequence 33, Appl
c	66	14.6	58.4	1059	2	US-08-867-087B-14	Sequence 14, Appl
c	67	14.6	58.4	1180	2	US-08-867-087B-16	Sequence 16, Appl
c	68	14.6	58.4	1269	1	US-08-844-064-5	Sequence 5, Appli
c	69	14.6	58.4	1269	3	US-09-009-433-5	Sequence 5, Appli
c	70	14.6	58.4	1635	1	US-08-133-347-1	Sequence 1, Appli
c	71	14.6	58.4	1635	1	US-08-133-347-3	Sequence 3, Appli
c	72	14.6	58.4	1735	4	US-09-163-444-10	Sequence 10, Appl
c	73	14.6	58.4	2162	2	US-08-870-518-5	Sequence 5, Appli
c	74	14.6	58.4	2487	3	US-08-481-435-1	Sequence 1, Appli
c	75	14.6	58.4	2652	2	US-08-953-492-1	Sequence 1, Appli
c	76	14.6	58.4	5484	4	US-09-632-580A-3	Sequence 3, Appli
c	77	14.6	58.4	5501	1	US-08-484-438-1	Sequence 1, Appli
c	78	14.6	58.4	5555	1	US-08-484-438-3	Sequence 3, Appli
c	79	14.6	58.4	14311	4	US-08-646-695-1	Sequence 1, Appli
	80	14.6	58.4	14311	4	US-08-646-695-7	Sequence 7, Appli
c	81	14.6	58.4	14311	5	PCT-US96-06053-1	Sequence 1, Appli
	82	14.6	58.4	14311	5	PCT-US96-06053-7	Sequence 7, Appli
c	83	14.6	58.4	50937	4	US-09-428-517-1	Sequence 1, Appli
c	84	14.4	57.6	293	2	US-08-630-822A-71	Sequence 71, Appl
c	85	14.4	57.6	293	2	US-09-005-069-71	Sequence 71, Appl
c	86	14.4	57.6	447	1	US-08-565-386-4	Sequence 4, Appli
	87	14.4	57.6	681	5	PCT-US94-04361-22	Sequence 22, Appl
c	88	14.4	57.6	852	4	US-09-586-935-6	Sequence 6, Appli
	89	14.4	57.6	960	3	US-09-032-372-10	Sequence 10, Appl
c	90	14.4	57.6	1309	1	US-07-926-788A-1	Sequence 1, Appli
	91	14.4	57.6	1322	1	US-08-468-709B-5	Sequence 5, Appli
	92	14.4	57.6	1322	2	US-08-241-664B-5	Sequence 5, Appli
	93	14.4	57.6	1322	5	PCT-US93-03936-5	Sequence 5, Appli
c	94	14.4	57.6	1381	4	US-09-426-557-5	Sequence 5, Appli
c	95	14.4	57.6	1432	1	US-07-914-282D-3	Sequence 3, Appli
c	96	14.4	57.6	1432	1	US-08-276-887A-3	Sequence 3, Appli
c	97	14.4	57.6	1432	5	PCT-US93-02460-3	Sequence 3, Appli
c	98	14.4	57.6	1463	4	US-09-426-557-1	Sequence 1, Appli
	99	14.4	57.6	1475	1	US-08-097-938-1	Sequence 1, Appli
	100	14.4	57.6	1475	1	US-08-476-000-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Enhanced Biotin Biosynthesis in Plant Tissue
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/351,970
;      FILING DATE:  08-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-401-068-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.0015;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||||||||||||||||||||||||
Db      99 GCATTTCGATCCTCGTCAGGTGCAG 123


RESULT   2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-846-338-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.0015;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||||||||||||||||||||||||
Db      99 GCATTTCGATCCTCGTCAGGTGCAG 123


RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A-15/9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1157
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 117
;      OTHER INFORMATION:  /product= "Biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioB"
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2295..3050
;      OTHER INFORMATION:  /codon_start= 2295
;      OTHER INFORMATION:  /function= "involved in pimeloyl-CoA synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /gene= "bioC"
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3750..5039
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3750
;      OTHER INFORMATION:  /EC_number= 2.6.1.62
;      OTHER INFORMATION:  /product= "DAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioA"
;      OTHER INFORMATION:  /number= 5
;      OTHER INFORMATION:  /standard_name=
;      OTHER INFORMATION:  "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
;      OTHER INFORMATION:  aminotransf."
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  5098..5574
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 5098
;      OTHER INFORMATION:  /function= "unknown, involved in biotin synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "ORFI"
;      OTHER INFORMATION:  /number= 6
;    FEATURE:
;      NAME/KEY:  -10_signal
;      LOCATION:  45..49
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  -35_signal
;      LOCATION:  23..28
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  105..119
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "bioB RBS no.9"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  2284..2297
;      OTHER INFORMATION:  /standard_name= "bioC RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3742..3752
;      OTHER INFORMATION:  /standard_name= "bioA RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  5088..5100
;      OTHER INFORMATION:  /standard_name= "ORFI RBS"
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  5583..5644
;      OTHER INFORMATION:  /standard_name= "rho-independent
;      OTHER INFORMATION:  transcriptional terminator"
;    FEATURE:
;      NAME/KEY:  stem_loop
;      LOCATION:  5583..5605
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..96
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "promoter ptac"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-1


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0022;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||||||||||||||||||||||||
Db     215 GCATTTCGATCCTCGTCAGGTGCAG 239


RESULT   4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A15-9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1154..2308
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1154
;      OTHER INFORMATION:  /EC_number= 2.3.1.47
;      OTHER INFORMATION:  /product= "KAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "8-Amino-7-oxononanoate synthase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3043..3753
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3043
;      OTHER INFORMATION:  /EC_number= 6.3.3.3
;      OTHER INFORMATION:  /product= "DTB synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioD"
;      OTHER INFORMATION:  /number= 4
;      OTHER INFORMATION:  /standard_name= "Dethiobiotin synthase"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1141..1156
;      OTHER INFORMATION:  /standard_name= "bioF RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3030..3045
;      OTHER INFORMATION:  /standard_name= "bioD RBS"
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-6


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0022;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||||||||||||||||||||||||
Db    215 GCATTTCGATCCTCGTCAGGTGCAG 239


RESULT   5
US-08-726-725-5
; Sequence 5, Application US/08726725
; Patent No. 5773290
;  GENERAL INFORMATION:
;    APPLICANT:  Gould, Michael N.
;    APPLICANT:  Chen, Kai-Shun
;    TITLE OF INVENTION:  MAMMARY GLAND-SPECIFIC PROMOTERS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  411 East Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  Wisconsin
;      COUNTRY:  U.S.A.
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/726,725
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baker, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  960296.93863
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5709
;      TELEFAX:  (414) 271-3552
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  876 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-726-725-5


  Query Match             68.8%;  Score 17.2;  DB 1;  Length 876;
  Best Local Similarity   86.4%;  Pred. No. 11;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgc 23
            || ||| ||||||||||||| ||
Db    303 CACTTCCATCCTCGTCAGGGGC 324


RESULT   6
US-08-317-880-3/c
; Sequence 3, Application US/08317880
; Patent No. 5593862
;  GENERAL INFORMATION:
;    APPLICANT:  Hall, Linda M.
;    TITLE OF INVENTION:  Proteins for Cation Channel Expression
;    TITLE OF INVENTION:  and Function
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/317,880
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  Rogalskyj, Peter
;      REGISTRATION NUMBER:  38,601
;      REFERENCE/DOCKET NUMBER:  19226/610 (R-5261)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1634
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1045 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-317-880-3


  Query Match             64.0%;  Score 16;  DB 1;  Length 1045;
  Best Local Similarity   79.2%;  Pred. No. 45;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           | |||| ||||| | || ||||||
Db     163 CCTTTCTATCCTTGCCACGTGCAG 140


RESULT   7
US-08-782-396-3/c
; Sequence 3, Application US/08782396
; Patent No. 5871940
;  GENERAL INFORMATION:
;    APPLICANT:  Hall, Linda M.
;    APPLICANT:  Feng, Guoping
;    TITLE OF INVENTION:  PROTEINS FOR CATION CHANNEL EXPRESSION
;    TITLE OF INVENTION:  AND FUNCTION
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/782,396
;      FILING DATE:  13-JAN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/317,880
;      FILING DATE:  04-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rogalskyj, Peter
;      REGISTRATION NUMBER:  38,601
;      REFERENCE/DOCKET NUMBER:  19226/611 (R-5261B)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1634
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1045 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-782-396-3


  Query Match             64.0%;  Score 16;  DB 2;  Length 1045;
  Best Local Similarity   79.2%;  Pred. No. 45;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           | |||| ||||| | || ||||||
Db     163 CCTTTCTATCCTTGCCACGTGCAG 140


RESULT   8
US-09-082-088-1/c
; Sequence 1, Application US/09082088
; Patent No. 6130067
;  GENERAL INFORMATION:
;    APPLICANT:  TSUI, PING
;    TITLE OF INVENTION:  HUMAN EDG3sb GENE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ratner & Prestia
;      STREET:  P.O. Box 980
;      CITY:  Valley Forge
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/082,088
;      FILING DATE:  20-MAY-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Prestia, Paul F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GP-70453
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1137 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-082-088-1


  Query Match             64.0%;  Score 16;  DB 3;  Length 1137;
  Best Local Similarity   79.2%;  Pred. No. 46;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           ||||| |||| | |||| |||| |
Db     429 CATTTTGATCATTGTCAAGTGCCG 406


RESULT   9
US-08-965-762-7/c
; Sequence 7, Application US/08965762
; Patent No. 6280963
; GENERAL INFORMATION:
;  APPLICANT: Koltin, Yigal
;  APPLICANT: Gavrias, Victoria
;  TITLE OF INVENTION: ESSENTIAL FUNGAL GENES AND THEIR USE
;  FILE REFERENCE: 07334/062001
;  CURRENT APPLICATION NUMBER: US/08/965,762
;  CURRENT FILING DATE: 1997-11-07
;  NUMBER OF SEQ ID NOS: 35
```

```
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 1792
;   TYPE: DNA
;   ORGANISM: Aspergillus nidulans
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (230)...(309)
;   NAME/KEY: CDS
;   LOCATION: (375)...(815)
;   NAME/KEY: CDS
;   LOCATION: (876)...(1149)
;   NAME/KEY: CDS
;   LOCATION: (1200)...(1475)
US-08-965-762-7


  Query Match             64.0%;  Score 16;  DB 4;  Length 1792;
  Best Local Similarity   79.2%;  Pred. No. 51;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgcag 25
          |||||||| | |   |||||||||
Db    708 CATTTCGAACATGCGCAGGTGCAG 685


RESULT  10
US-08-330-537-3/c
; Sequence 3, Application US/08330537
; Patent No. 5663048
;  GENERAL INFORMATION:
;    APPLICANT:  WINKFEIN, ROBERT J.
;    APPLICANT:  DIXON, GORDON H.
;    TITLE OF INVENTION:  Y-CHROMOSOME SPECIFIC POLYNUCLEOTIDE
;    TITLE OF INVENTION:  PROBES FOR PRENATAL SEXING
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  699 PRINCE, P.O. BOX 1404
;      CITY:  ALEXANDRIA
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/330,537
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/030,485
;      FILING DATE:  13-MAY-1993
;      APPLICATION NUMBER:  PCT/CA91/00353
;      FILING DATE:  03-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CA 2,026,926
;      FILING DATE:  04-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DILLAHUNTY, T. GENE
;      REGISTRATION NUMBER:  25,423
;      REFERENCE/DOCKET NUMBER:  028722-080
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-854-7400
;      TELEFAX:  415-854-8275
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3713 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-330-537-3


  Query Match             64.0%;  Score 16;  DB 1;  Length 3713;
  Best Local Similarity   79.2%;  Pred. No. 59;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgca 24
          |||  ||| |||| |||||| |||
Db   1537 GCAGATCGGTCCTTGTCAGGAGCA 1514


RESULT  11
US-08-330-537-4/c
; Sequence 4, Application US/08330537
; Patent No. 5663048
;  GENERAL INFORMATION:
;    APPLICANT:  WINKFEIN, ROBERT J.
;    APPLICANT:  DIXON, GORDON H.
;    TITLE OF INVENTION:  Y-CHROMOSOME SPECIFIC POLYNUCLEOTIDE
;    TITLE OF INVENTION:  PROBES FOR PRENATAL SEXING
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  699 PRINCE, P.O. BOX 1404
;      CITY:  ALEXANDRIA
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/330,537
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/030,485
;      FILING DATE:  13-MAY-1993
;      APPLICATION NUMBER:  PCT/CA91/00353
;      FILING DATE:  03-OCT-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CA 2,026,926
;      FILING DATE:  04-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DILLAHUNTY, T. GENE
;      REGISTRATION NUMBER:  25,423
;      REFERENCE/DOCKET NUMBER:  028722-080
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-854-7400
;      TELEFAX:  415-854-8275
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3713 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-330-537-4


  Query Match             64.0%;  Score 16;  DB 1;  Length 3713;
  Best Local Similarity   79.2%;  Pred. No. 59;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgca 24
          |||  ||| |||| |||||| |||
Db   1537 GCAGATCGGTCCTTGTCAGGAGCA 1514


RESULT  12
US-08-317-880-1/c
```

```
; Sequence 1, Application US/08317880
; Patent No. 5593862
;  GENERAL INFORMATION:
;    APPLICANT:  Hall, Linda M.
;    TITLE OF INVENTION:  Proteins for Cation Channel Expression
;    TITLE OF INVENTION:  and Function
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/317,880
;      FILING DATE:  04-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rogalskyj, Peter
;      REGISTRATION NUMBER:  38,601
;      REFERENCE/DOCKET NUMBER:  19226/610 (R-5261)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1634
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3954 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-317-880-1


  Query Match             64.0%;  Score 16;  DB 1;  Length 3954;
  Best Local Similarity   79.2%;  Pred. No. 60;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           | |||| ||||| | || ||||||
Db     163 CCTTTCTATCCTTGCCACGTGCAG 140


RESULT  13
US-08-782-396-1/c
; Sequence 1, Application US/08782396
; Patent No. 5871940
;  GENERAL INFORMATION:
;    APPLICANT:  Hall, Linda M.
;    APPLICANT:  Feng, Guoping
;    TITLE OF INVENTION:  PROTEINS FOR CATION CHANNEL EXPRESSION
;    TITLE OF INVENTION:  AND FUNCTION
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/782,396
;      FILING DATE:  13-JAN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/317,880
;      FILING DATE:  04-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rogalskyj, Peter
;      REGISTRATION NUMBER:  38,601
;      REFERENCE/DOCKET NUMBER:  19226/611 (R-5261B)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1634
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3954 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-782-396-1


  Query Match             64.0%;  Score 16;  DB 2;  Length 3954;
  Best Local Similarity   79.2%;  Pred. No. 60;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           | |||| ||||| | || ||||||
Db     163 CCTTTCTATCCTTGCCACGTGCAG 140


RESULT  14
US-08-399-561-1/c
; Sequence 1, Application US/08399561
; Patent No. 5792903
;  GENERAL INFORMATION:
;    APPLICANT:  Hirschberg, Joseph
;    APPLICANT:  Cunningham Jr., Francis X.
;    APPLICANT:  Gantt, Elisabeth
;    TITLE OF INVENTION:  Lycopene Cyclase Gene
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kohn & Associates
;      STREET:  30500 No. 5792903thwestern Highway, Suite 410
;      CITY:  Farmington Hills
;      STATE:  MI
;      COUNTRY:  US
;      ZIP:  48334
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/399,561
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kohn, Kenneth I.
;      REGISTRATION NUMBER:  30,955
;      REFERENCE/DOCKET NUMBER:  P-305 (Hebrew Univ.)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  810-539-5050
;      TELEFAX:  810-539-5055
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4928 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
```

```
;      LOCATION:  2029..3261
US-08-399-561-1


  Query Match             64.0%;  Score 16;  DB 1;  Length 4928;
  Best Local Similarity   79.2%;  Pred. No. 62;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           ||  || ||| |||||||||| |||
Db    4907 CAGCTCCATCATCGTCAGGTTCAG 4884


RESULT  15
US-08-875-533-59/c
; Sequence 59, Application US/08875533
; Patent No. 6254870
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  No. 6254870el c-MPL Ligands
;    NUMBER OF SEQUENCES:  73
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/875,533
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/383,035
;      FILING DATE:  04-FEB-1995
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  996 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-875-533-59


  Query Match             62.4%;  Score 15.6;  DB 4;  Length 996;
  Best Local Similarity   81.8%;  Pred. No. 71;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtg 22
           |||||  |||||||||  || ||
Db     377 GCATTGGGATCCTCGTGAGCTG 356


RESULT  16
US-08-933-750C-71/c
; Sequence 71, Application US/08933750C
; Patent No. 5932442
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  HUMAN REGULATORY MOLECULES
;    NUMBER OF SEQUENCES:  98
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/933,750C
;      FILING DATE:  September 23, 1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0356 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  71:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1139 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  LATRTUT02
;      CLONE:  1404643
US-08-933-750C-71


  Query Match             62.4%;  Score 15.6;  DB 2;  Length 1139;
  Best Local Similarity   81.8%;  Pred. No. 73;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 atttcgatcctcgtcaggtgca 24
           |||| |||||||  || |||| |
Db     668 ATTTGGATCCTCATCTGGTGTA 647


RESULT  17
US-09-234-613-71/c
; Sequence 71, Application US/09234613
; Patent No. 6132973
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  HUMAN REGULATORY MOLECULES
;    NUMBER OF SEQUENCES:  98
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/234,613
;      FILING DATE:
;      CLASSIFICATION:
```

```
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/933,750
;       FILING DATE:  September 23, 1997
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Billings, Lucy J.
;       REGISTRATION NUMBER:  36,749
;       REFERENCE/DOCKET NUMBER:  PF-0356 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-855-0555
;       TELEFAX:  415-845-4166
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  71:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1139 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;       LIBRARY:  LATRTUT02
;       CLONE:  1404643
US-09-234-613-71


  Query Match             62.4%;  Score 15.6;  DB 3;  Length 1139;
  Best Local Similarity   81.8%;  Pred. No. 73;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 atttcgatcctcgtcaggtgca 24
           |||| ||||||| || |||| |
Db     668 ATTTGGATCCTCATCTGGTGTA 647


RESULT  18
US-07-841-646-8
; Sequence 8, Application US/07841646
; Patent No. 5266683
;  GENERAL INFORMATION:
;     APPLICANT:  OPPERMANN, HERMANN
;     APPLICANT:  OZKAYNAK, ENGIN
;     APPLICANT:  KUBERASAMPATH, THANGAVEL
;     APPLICANT:  RUEGER, DAVID C.
;     APPLICANT:  PANG, ROY H.L.
;     TITLE OF INVENTION:  OSTEOGENIC DEVICES
;     NUMBER OF SEQUENCES:  33
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;       STREET:  53 STATE STREET
;       CITY:  BOSTON
;       STATE:  MASSACHUSETTS
;       COUNTRY:  U.S.A.
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/841,646
;       FILING DATE:  19920221
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 810,560
;       FILING DATE:  20-DEC-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 827,052
;       FILING DATE:  28-JAN-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 660,162
;       FILING DATE:  22-FEB-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 621,988
;       FILING DATE:  04-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 621,849
;       FILING DATE:  04-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 616,374
;       FILING DATE:  21-NOV-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 600,024
;       FILING DATE:  18-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 599,543
;       FILING DATE:  18-OCT-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 579,865
;       FILING DATE:  07-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 569,920
;       FILING DATE:  20-AUG-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 483,913
;       FILING DATE:  22-FEB-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 422,613
;       FILING DATE:  17-OCT-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 315,342
;       FILING DATE:  23-FEB-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 232,630
;       FILING DATE:  15-AUG-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 179,460
;       FILING DATE:  08-APR-1988
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER, EDMUND R.
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  CRP-001CP6
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7000
;       TELEFAX:  617/248-7100
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  516 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  YES
;     ANTI-SENSE:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..507
;       OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;       OTHER INFORMATION:  /product= "OP1A"
;       OTHER INFORMATION:  /note= "OP1A FUSION"
US-07-841-646-8


  Query Match             61.6%;  Score 15.4;  DB 1;  Length 516;
  Best Local Similarity   76.0%;  Pred. No. 78;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           | | |||||||||| ||| ||||
Db     174 GGAATTCGATCCTCACCAGAGGCAG 198


RESULT  19
US-08-147-023-8
; Sequence 8, Application US/08147023
; Patent No. 5468845
;  GENERAL INFORMATION:
;     APPLICANT:  OPPERMANN, HERMANN
;     APPLICANT:  OZKAYNAK, ENGIN
```

```
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/147,023
;      FILING DATE:  21-FEB-1992
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 810,560
;      FILING DATE:  20-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 827,052
;      FILING DATE:  28-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 660,162
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,988
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,849
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 616,374
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 600,024
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 599,543
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 579,865
;      FILING DATE:  07-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 569,920
;      FILING DATE:  20-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 483,913
;      FILING DATE:  22-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 422,613
;      FILING DATE:  17-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 315,342
;      FILING DATE:  23-FEB-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 232,630
;      FILING DATE:  15-AUG-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 179,460
;      FILING DATE:  08-APR-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-001CP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  516 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..507
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1A"
;      OTHER INFORMATION:  /note= "OP1A FUSION"
US-08-147-023-8


  Query Match             61.6%;  Score 15.4;  DB 1;  Length 516;
  Best Local Similarity   76.0%;  Pred. No. 78;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           | | ||||||||||  |||  ||||
Db     174 GGAATTCGATCCTCACCAGAGGCAG 198


RESULT  20
US-08-447-570-8
; Sequence 8, Application US/08447570
; Patent No. 5714589
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/447,570
;      FILING DATE:  21-FEB-1992
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 810,560
;      FILING DATE:  20-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 827,052
;      FILING DATE:  28-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 660,162
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,988
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,849
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 616,374
;      FILING DATE:  21-NOV-1990
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 600,024
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 599,543
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 579,865
;      FILING DATE:  07-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 569,920
;      FILING DATE:  20-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 483,913
;      FILING DATE:  22-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 422,613
;      FILING DATE:  17-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 315,342
;      FILING DATE:  23-FEB-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 232,630
;      FILING DATE:  15-AUG-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 179,460
;      FILING DATE:  08-APR-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-001CP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  516 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..507
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1A"
;      OTHER INFORMATION:  /note= "OP1A FUSION"
US-08-447-570-8


  Query Match             61.6%;  Score 15.4;  DB 1;  Length 516;
  Best Local Similarity   76.0%;  Pred. No. 78;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           | | ||||||||||  |||  ||||
Db     174 GGAATTCGATCCTCACCAGAGGCAG 198


RESULT  21
US-08-449-700-8
; Sequence 8, Application US/08449700
; Patent No. 5863758
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  OSTEOGENIC DEVICES
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/449,700
;      FILING DATE:  21-FEB-1992
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 810,560
;      FILING DATE:  20-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 827,052
;      FILING DATE:  28-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 660,162
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,988
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 621,849
;      FILING DATE:  04-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 616,374
;      FILING DATE:  21-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 600,024
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 599,543
;      FILING DATE:  18-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 579,865
;      FILING DATE:  07-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 569,920
;      FILING DATE:  20-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 483,913
;      FILING DATE:  22-FEB-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 422,613
;      FILING DATE:  17-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 315,342
;      FILING DATE:  23-FEB-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 232,630
;      FILING DATE:  15-AUG-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 179,460
;      FILING DATE:  08-APR-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  CRP-001CP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  516 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..507
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1A"
;      OTHER INFORMATION:  /note= "OP1A FUSION"
US-08-449-700-8


  Query Match             61.6%;  Score 15.4;  DB 2;  Length 516;
  Best Local Similarity   76.0%;  Pred. No. 78;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          | | ||||||||||  |||  ||||
Db    174 GGAATTCGATCCTCACCAGAGGCAG 198


RESULT  22
US-08-449-699A-8
; Sequence 8, Application US/08449699A
; Patent No. 5958441
;  GENERAL INFORMATION:
;    APPLICANT:  OPPERMANN, HERMANN
;    APPLICANT:  OZKAYNAK, ENGIN
;    APPLICANT:  KUBERASAMPATH, THANGAVEL
;    APPLICANT:  RUEGER, DAVID C.
;    APPLICANT:  PANG, ROY H.L.
;    TITLE OF INVENTION:  ANTIBODIES TO OSTEOGENIC PROTEINS
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  125 HIGH STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  U.S.A.
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/449,699A
;      FILING DATE:  24-MAY-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/147,023
;      FILING DATE:  01-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER, EDMUND R.
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  STK-001CP6CN
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  516 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..507
;      OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
;      OTHER INFORMATION:  /product= "OP1A"
;      OTHER INFORMATION:  /note= "OP1A FUSION"
US-08-449-699A-8


  Query Match             61.6%;  Score 15.4;  DB 2;  Length 516;
  Best Local Similarity   76.0%;  Pred. No. 78;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          | | ||||||||||  |||  ||||
Db    174 GGAATTCGATCCTCACCAGAGGCAG 198


RESULT  23
US-09-193-377B-2
; Sequence 2, Application US/09193377B
; Patent No. 6221594
; GENERAL INFORMATION:
;  APPLICANT:  Burrell, Paul
;  APPLICANT:  Blackall, Linda
;  APPLICANT:  Keller, Jurg
;  TITLE OF INVENTION:  METHOD FOR THE DETECTION OF AQUATIC
;  TITLE OF INVENTION:  NITRITE OXIDISING MICROORGANISMS OF THE GENUS NITROSPIRA
;  FILE REFERENCE:  CULLN20.001AUS
;  CURRENT APPLICATION NUMBER: US/09/193,377B
;  CURRENT FILING DATE:  1998-11-17
;  NUMBER OF SEQ ID NOS:  62
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1407
;   TYPE: DNA
;   ORGANISM: Nitrospira
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1407)
;   OTHER INFORMATION: n = A,T,C or G
US-09-193-377B-2


  Query Match             61.6%;  Score 15.4;  DB 4;  Length 1407;
  Best Local Similarity   76.0%;  Pred. No. 96;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          |||  | | |||||  |||||||| |
Db    976 gcaataccatcctgctcaggtgctg 1000


RESULT  24
US-09-193-377B-4
; Sequence 4, Application US/09193377B
; Patent No. 6221594
; GENERAL INFORMATION:
;  APPLICANT:  Burrell, Paul
;  APPLICANT:  Blackall, Linda
;  APPLICANT:  Keller, Jurg
;  TITLE OF INVENTION:  METHOD FOR THE DETECTION OF AQUATIC
;  TITLE OF INVENTION:  NITRITE OXIDISING MICROORGANISMS OF THE GENUS NITROSPIRA
;  FILE REFERENCE:  CULLN20.001AUS
;  CURRENT APPLICATION NUMBER: US/09/193,377B
;  CURRENT FILING DATE:  1998-11-17
;  NUMBER OF SEQ ID NOS:  62
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 1420
;   TYPE: DNA
;   ORGANISM: Nitrospira
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1420)
;   OTHER INFORMATION: n = A,T,C or G
US-09-193-377B-4
```

```
  Query Match             61.6%;  Score 15.4;  DB 4;  Length 1420;
  Best Local Similarity   76.0%;  Pred. No. 96;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          ||| | | |||||  |||||||| |
Db    958 gcaataccatcctgctcaggtgctg 982


RESULT  25
US-09-193-377B-1
; Sequence 1, Application US/09193377B
; Patent No. 6221594
; GENERAL INFORMATION:
;  APPLICANT:  Burrell, Paul
;  APPLICANT:  Blackall, Linda
;  APPLICANT:  Keller, Jurg
;  TITLE OF INVENTION:  METHOD FOR THE DETECTION OF AQUATIC
;  TITLE OF INVENTION:  NITRITE OXIDISING MICROORGANISMS OF THE GENUS NITROSPIRA
;  FILE REFERENCE:  CULLN20.001AUS
;  CURRENT APPLICATION NUMBER: US/09/193,377B
;  CURRENT FILING DATE:  1998-11-17
;  NUMBER OF SEQ ID NOS:  62
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1428
;   TYPE: DNA
;   ORGANISM: Nitrospira
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1428)
;   OTHER INFORMATION: n = A,T,C or G
US-09-193-377B-1


  Query Match             61.6%;  Score 15.4;  DB 4;  Length 1428;
  Best Local Similarity   76.0%;  Pred. No. 96;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          ||| | | |||||  |||||||| |
Db    967 gcaataccatcctgctcaggtgctg 991


RESULT  26
US-08-199-776-1/c
; Sequence 1, Application US/08199776
; Patent No. 5594120
;  GENERAL INFORMATION:
;    APPLICANT:  Brenner, Michael B.
;    APPLICANT:  Parker, Christina M.
;    TITLE OF INVENTION:  No. 5594120el integrin alpha subunit
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield and Sacks, P.C.
;      STREET:  600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/199,776
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plumer, Elizabeth R.
;      REGISTRATION NUMBER:  36,637
;      REFERENCE/DOCKET NUMBER:  B0801/7020
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-720-3500
;      TELEFAX:  617-720-2441
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3933 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL TYPE:  mucosal lymphocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  126..3662
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  180..3659
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  126..179
US-08-199-776-1


  Query Match             61.6%;  Score 15.4;  DB 1;  Length 3933;
  Best Local Similarity   76.0%;  Pred. No. 1.2e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          || |||  |||||| |||  |||||
Db   2778 GCTTTTGCATCCTCTTCAACTGCAG 2754


RESULT  27
US-08-663-731-1/c
; Sequence 1, Application US/08663731
; Patent No. 6057423
;  GENERAL INFORMATION:
;    APPLICANT:  Brenner, Michael B.
;    APPLICANT:  Parker, Christina M.
;    TITLE OF INVENTION:  No. 6057423el integrin alpha subunit
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield and Sacks, P.C.
;      STREET:  600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/663,731
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/199,776
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plumer, Elizabeth R.
;      REGISTRATION NUMBER:  36,637
;      REFERENCE/DOCKET NUMBER:  B0801/7020
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-720-3500
;      TELEFAX:  617-720-2441
```

```
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3933 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL TYPE:  mucosal lymphocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  126..3662
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  180..3659
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  126..179
US-08-663-731-1


  Query Match             61.6%;  Score 15.4;  DB 3;  Length 3933;
  Best Local Similarity   76.0%;  Pred. No. 1.2e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          || |||  |||||| |||  |||||
Db   2778 GCTTTTGCATCCTCTTCAACTGCAG 2754


RESULT  28
US-08-879-338-1/c
; Sequence 1, Application US/08879338A
; Patent No. 6063906
; GENERAL INFORMATION:
;  APPLICANT: Brenner, Michael B.
;  APPLICANT: Parker, Christina M.
;  TITLE OF INVENTION: Antibodies to No. 6063906el Integrin Alpha
;  TITLE OF INVENTION: Subunit
;  FILE REFERENCE: B0801/7080/ERP
;  CURRENT APPLICATION NUMBER: US/08/879,338A
;  CURRENT FILING DATE: 1997-06-20
;  EARLIER APPLICATION NUMBER: US 08/663,731
;  EARLIER FILING DATE: 1996-06-14
;  EARLIER APPLICATION NUMBER: US 08/199,776
;  EARLIER FILING DATE: 1994-02-18
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 3933
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (126)...(3659)
;   NAME/KEY: mat_peptide
;   LOCATION: (180)...(3659)
;   NAME/KEY: sig_peptide
;   LOCATION: (126)...(179)
US-08-879-338-1


  Query Match             61.6%;  Score 15.4;  DB 3;  Length 3933;
  Best Local Similarity   76.0%;  Pred. No. 1.2e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          || |||  |||||| |||  |||||
Db   2778 GCTTTTGCATCCTCTTCAACTGCAG 2754


RESULT  29
PCT-US95-02044-1/c
; Sequence 1, Application PC/TUS9502044
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Novel integrin alpha subunit
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield and Sacks, P.C.
;      STREET:  600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/02044
;      FILING DATE:  herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/199,776
;      FILING DATE:  18 February 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plumer, Elizabeth R.
;      REGISTRATION NUMBER:  36,637
;      REFERENCE/DOCKET NUMBER:  B0801/7020
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-720-3500
;      TELEFAX:  617-720-2441
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3933 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL TYPE:  mucosal lymphocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  126..3662
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  180..3659
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  126..179
PCT-US95-02044-1


  Query Match             61.6%;  Score 15.4;  DB 5;  Length 3933;
  Best Local Similarity   76.0%;  Pred. No. 1.2e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          || |||  |||||| |||  |||||
Db   2778 GCTTTTGCATCCTCTTCAACTGCAG 2754


RESULT  30
US-09-051-019-1
; Sequence 1, Application US/09051019
; Patent No. 6103229
;  GENERAL INFORMATION:
;    APPLICANT:  KAHMANN, Regine and QUADBECK-SEEGER, Claudia
```

```
;    TITLE OF INVENTION:  Regulatory gene from Ustilago maydis
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Keil & Weinkauf
;      STREET:  1101 Connecticut Avenue
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.2 Mb storage
;      COMPUTER:  IBM AT-compatible, Pentium processor
;      OPERATING SYSTEM:  Windows 98
;      SOFTWARE:  WordPerfect version 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/051,019
;      FILING DATE:  31-MAR-1998
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8931 base pairs
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-051-019-1


  Query Match             61.6%;  Score 15.4;  DB 3;  Length 8931;
  Best Local Similarity   76.0%;  Pred. No. 1.4e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtgcag 25
          ||  |||  || ||||||| ||||
Db   6146 GCCATTCCTTCGTCGTCAGGGGCAG 6170


RESULT  31
US-08-441-971-16/c
; Sequence 16, Application US/08441971
; Patent No. 6071693
;  GENERAL INFORMATION:
;    APPLICANT:  Tai-An Cha
;    TITLE OF INVENTION:  HCV GENOMIC SEQUENCES FOR
;    TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;    NUMBER OF SEQUENCES:  147
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
;      STREET:  600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  MS-DOS Version 3.3
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/441,971
;      FILING DATE:  16-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/221,653
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/881,528
;      FILING DATE:
;      APPLICATION NUMBER:  07/697,326
;      FILING DATE:  8 May 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Janiuk, Anthony J.
;      REGISTRATION NUMBER:  29,809
;      REFERENCE/DOCKET NUMBER:  C0772/7000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 720-3500
;      TELEFAX:  (617) 720-2441
;      TELEX:  EZEKIEL
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  340 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  ns5arg6
US-08-441-971-16


  Query Match             60.8%;  Score 15.2;  DB 3;  Length 340;
  Best Local Similarity   85.0%;  Pred. No. 90;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcagg 20
          ||| ||||| |||| |||||
Db     81 GCAGTTCGAGCCTCCTCAGG 62


RESULT  32
US-08-221-653-16/c
; Sequence 16, Application US/08221653
; Patent No. 6190864
;  GENERAL INFORMATION:
;    APPLICANT:  Tai-An Cha
;    TITLE OF INVENTION:  HCV GENOMIC SEQUENCES FOR
;    TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;    NUMBER OF SEQUENCES:  147
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
;      STREET:  600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  MS-DOS Version 3.3
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/221,653
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/881,528
;      FILING DATE:
;      APPLICATION NUMBER:  07/697,326
;      FILING DATE:  8 May 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Janiuk, Anthony J.
;      REGISTRATION NUMBER:  29,809
;      REFERENCE/DOCKET NUMBER:  C0772/7000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 720-3500
;      TELEFAX:  (617) 720-2441
;      TELEX:  EZEKIEL
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  340 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  ns5arg6
US-08-221-653-16
```

```
  Query Match             60.8%;  Score 15.2;  DB 4;  Length 340;
  Best Local Similarity   85.0%;  Pred. No. 90;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcagg 20
          ||| ||||| |||| |||||
Db     81 GCAGTTCGAGCCTCCTCAGG 62


RESULT  33
US-08-442-144A-16/c
; Sequence 16, Application US/08442144A
; Patent No. 6214583
;  GENERAL INFORMATION:
;    APPLICANT:  Tai-An Cha
;    APPLICANT:  Eileen Beall
;    APPLICANT:  Bruce Irvine
;    APPLICANT:  Janice Kolberg
;    APPLICANT:  Michael S. Urdea
;    TITLE OF INVENTION:  HCV GENOMIC SEQUENCES FOR
;    TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;    NUMBER OF SEQUENCES:  148
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 Inch
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows NT
;      SOFTWARE:  Microsoft Word 97
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/442,144A
;      FILING DATE:  MAY 16, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/221,653
;      FILING DATE:  APRIL 1, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Doreen Yatko Trujillo
;      REGISTRATION NUMBER:  35,719
;      REFERENCE/DOCKET NUMBER:  CHIR-0121
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  340 Nucleotides
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    MOLECULE TYPE:  DNA
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  ns5arg6
US-08-442-144A-16


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 340;
  Best Local Similarity   85.0%;  Pred. No. 90;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcagg 20
          ||| ||||| |||| |||||
Db     81 GCAGTTCGAGCCTCCTCAGG 62


RESULT  34
US-08-441-970-16/c
; Sequence 16, Application US/08441970
; Patent No. 6297370
;  GENERAL INFORMATION:
;    APPLICANT:  Tai-An Cha
;    TITLE OF INVENTION:  HCV GENOMIC SEQUENCES FOR
;    TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;    NUMBER OF SEQUENCES:  147
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
;      STREET:  600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  MS-DOS Version 3.3
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/441,970
;      FILING DATE:  16-MAY-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/881,528
;      FILING DATE:  08-MAY-1992
;      APPLICATION NUMBER:  07/697,326
;      FILING DATE:  8 May 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Janiuk, Anthony J.
;      REGISTRATION NUMBER:  29,809
;      REFERENCE/DOCKET NUMBER:  C0772/7000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 720-3500
;      TELEFAX:  (617) 720-2441
;      TELEX:  EZEKIEL
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  340 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  ns5arg6
US-08-441-970-16


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 340;
  Best Local Similarity   85.0%;  Pred. No. 90;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcagg 20
          ||| ||||| |||| |||||
Db     81 GCAGTTCGAGCCTCCTCAGG 62


RESULT  35
US-08-528-199-2/c
; Sequence 2, Application US/08528199
; Patent No. 5763228
;  GENERAL INFORMATION:
;    APPLICANT:  KUBOTA, Michio
;    APPLICANT:  TSUSAKI, Keiji
;    APPLICANT:  SUGIMOTO, Toshiyuki
;    TITLE OF INVENTION:  RECOMBINANT ENZYME FOR CONVERTING
;    TITLE OF INVENTION:  MALTOSE INTO TREHALOSE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
```

```
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/528,199
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/485,126
;      FILING DATE:  14-SEP-1995
;      APPLICATION NUMBER:  US 08/485,126
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 156399/1994
;      FILING DATE:  16-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  KUBOTA=6A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1704 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-528-199-2


  Query Match             60.8%;  Score 15.2;  DB 1;  Length 1704;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       6 tcgatcctcgtcaggtgcag 25
           | |||| ||| |||||||||
Db    1295 TGGATCGTCGGCAGGTGCAG 1276


RESULT  36
US-08-528-199-5/c
; Sequence 5, Application US/08528199
; Patent No. 5763228
;  GENERAL INFORMATION:
;    APPLICANT:  KUBOTA, Michio
;    APPLICANT:  TSUSAKI, Keiji
;    APPLICANT:  SUGIMOTO, Toshiyuki
;    TITLE OF INVENTION:  RECOMBINANT ENZYME FOR CONVERTING
;    TITLE OF INVENTION:  MALTOSE INTO TREHALOSE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/528,199
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/485,126
;      FILING DATE:  14-SEP-1995
;      APPLICATION NUMBER:  US 08/485,126
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 156399/1994
;      FILING DATE:  16-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  KUBOTA=6A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1704 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1704
US-08-528-199-5


  Query Match             60.8%;  Score 15.2;  DB 1;  Length 1704;
  Best Local Similarity   85.0%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       6 tcgatcctcgtcaggtgcag 25
           | |||| ||| |||||||||
Db    1295 TGGATCGTCGGCAGGTGCAG 1276


RESULT  37
US-08-758-306-140/c
; Sequence 140, Application US/08758306
; Patent No. 5807743
;  GENERAL INFORMATION:
;    APPLICANT:  Stinchcomb, Dan T.
;    APPLICANT:  McSwiggen, James A.
;    TITLE OF INVENTION:  METHOD AND REAGENT FOR THE
;    TITLE OF INVENTION:  TREATMENT OF DISEASES
;    TITLE OF INVENTION:  ASSOCIATED WITH
;    TITLE OF INVENTION:  INTERLEUKIN-2 RECEPTOR
;    TITLE OF INVENTION:  GAMMA-CHAIN EXPRESSION
;    NUMBER OF SEQUENCES:  1379
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/758,306
;      FILING DATE:  December 3, 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Warburg, Richard J.
;       REGISTRATION NUMBER:  32,327
;       REFERENCE/DOCKET NUMBER:  212/132
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  140:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  27 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       OTHER INFORMATION:  The letter "N" stands for the stem II
;       OTHER INFORMATION:  region of a HH ribozyme.
US-08-758-306-140


  Query Match             60.0%;  Score 15;  DB 1;  Length 27;
  Best Local Similarity   75.0%;  Pred. No. 68;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           |||||   || || ||||||  |||
Db      26 CATTTGTTTCNTCATCAGGTTCAG 3


RESULT  38
US-08-758-306-716/c
; Sequence 716, Application US/08758306
; Patent No. 5807743
;  GENERAL INFORMATION:
;    APPLICANT:  Stinchcomb, Dan T.
;    APPLICANT:  McSwiggen, James A.
;    TITLE OF INVENTION:  METHOD AND REAGENT FOR THE
;    TITLE OF INVENTION:  TREATMENT OF DISEASES
;    TITLE OF INVENTION:  ASSOCIATED WITH
;    TITLE OF INVENTION:  INTERLEUKIN-2 RECEPTOR
;    TITLE OF INVENTION:  GAMMA-CHAIN EXPRESSION
;    NUMBER OF SEQUENCES:  1379
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/758,306
;      FILING DATE:  December 3, 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  212/132
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  716:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      OTHER INFORMATION:  The letter "N" stands for the stem II
;      OTHER INFORMATION:  region of a HH ribozyme.
US-08-758-306-716


  Query Match             60.0%;  Score 15;  DB 1;  Length 27;
  Best Local Similarity   75.0%;  Pred. No. 68;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           |||||   || || ||||||  |||
Db      26 CATTTGTTTCNTCATCAGGTCCAG 3


RESULT  39
US-08-758-306-1112/c
; Sequence 1112, Application US/08758306
; Patent No. 5807743
;  GENERAL INFORMATION:
;    APPLICANT:  Stinchcomb, Dan T.
;    APPLICANT:  McSwiggen, James A.
;    TITLE OF INVENTION:  METHOD AND REAGENT FOR THE
;    TITLE OF INVENTION:  TREATMENT OF DISEASES
;    TITLE OF INVENTION:  ASSOCIATED WITH
;    TITLE OF INVENTION:  INTERLEUKIN-2 RECEPTOR
;    TITLE OF INVENTION:  GAMMA-CHAIN EXPRESSION
;    NUMBER OF SEQUENCES:  1379
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/758,306
;      FILING DATE:  December 3, 1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  212/132
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  1112:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      OTHER INFORMATION:  The letter "N" stands for the stem II
;      OTHER INFORMATION:  region of a HH ribozyme.
US-08-758-306-1112


  Query Match             60.0%;  Score 15;  DB 1;  Length 27;
```

```
  Best Local Similarity   75.0%;  Pred. No. 68;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           |||||   || || ||||||| |||
Db      26 CATTTGTTTCNTCATCAGGTCCAG 3


RESULT  40
US-09-060-756-714
; Sequence 714, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
;  APPLICANT: Cole, Stewart
;  APPLICANT: Buchrieser-Brosch, Roland
;  APPLICANT: Gordon, Stephen
;  APPLICANT: Billault, Alain
;  TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
;  TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
;  TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
;  FILE REFERENCE: 3495-0169
;  CURRENT APPLICATION NUMBER: US/09/060,756
;  CURRENT FILING DATE: 1998-04-16
;  NUMBER OF SEQ ID NOS: 743
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 714
;   LENGTH: 233
;   TYPE: DNA
;   ORGANISM: Mycobacterium bovis
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (various positions within the sequence)
;   OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-714


  Query Match             60.0%;  Score 15;  DB 4;  Length 233;
  Best Local Similarity   78.3%;  Pred. No. 1.1e+02;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgc 23
           |||  ||| ||  ||||||||||
Db      71 gcagctcggtcagcgtcaggtgc 93


RESULT  41
US-08-998-416-995/c
; Sequence 995, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  995:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  666 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1611RP
US-08-998-416-995


  Query Match             60.0%;  Score 15;  DB 4;  Length 666;
  Best Local Similarity   78.3%;  Pred. No. 1.3e+02;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgca 24
           | ||  ||||  ||||||||||||
Db     167 CCTTGAGATCGCCGTCAGGTGCA 145


RESULT  42
US-08-998-416-1093/c
; Sequence 1093, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  1093:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  707 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1667RP
US-08-998-416-1093


  Query Match             60.0%;  Score 15;  DB 4;  Length 707;
  Best Local Similarity   78.3%;  Pred. No. 1.3e+02;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgca 24
           | ||  |||| ||||||||||||
Db     167 CCTTGAGATCGCCGTCAGGTGCA 145


RESULT  43
US-08-557-128-3/c
; Sequence 3, Application US/08557128
; Patent No. 5849524
;  GENERAL INFORMATION:
;    APPLICANT:  KONDO, Keiji
;    APPLICANT:  KAJIWARA, Susumu
;    APPLICANT:  MISAWA, No. 5849524ihiko
;    TITLE OF INVENTION:  TRANSFORMATION SYSTEMS FOR THE YEAST
;    TITLE OF INVENTION:  CANDIDA UTILIS AND THE EXPRESSION OF HETEROLOGOUS GENES
;    TITLE OF INVENTION:  THEREWITH
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/557,128
;      FILING DATE:  25-JAN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/JP95/01005
;      FILING DATE:  25-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 7-129287
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-285823
;      FILING DATE:  26-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 6-135015
;      FILING DATE:  25-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  49441/108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2330 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1259..2059
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1259..2059
US-08-557-128-3


  Query Match             60.0%;  Score 15;  DB 2;  Length 2330;
  Best Local Similarity   78.3%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgca 24
           |||| | |||| |||| |||| |
Db     121 CATTACTATCCACGTCGGGTGTA 99


RESULT  44
US-09-242-690A-34/c
; Sequence 34, Application US/09242690A
; Patent No. 6284534
; GENERAL INFORMATION:
;  APPLICANT: KONDO, KEIJI
;  APPLICANT:  MIURA, YUTAKA
;  TITLE OF INVENTION: YEAST VECTOR AND METHOD OF PRODUCING PROTEINS USING THE
;  TITLE OF INVENTION:  SAME
;  FILE REFERENCE: 049441/0118
;  CURRENT APPLICATION NUMBER: US/09/242,690A
;  CURRENT FILING DATE:  1999-02-23
;  PRIOR APPLICATION NUMBER: PCT/JP97/02924
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: JP 8/241062
;  PRIOR FILING DATE: 1996-08-23
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
;   LENGTH: 2330
;   TYPE: DNA
;   ORGANISM: Candida utilis
US-09-242-690A-34


  Query Match             60.0%;  Score 15;  DB 4;  Length 2330;
  Best Local Similarity   78.3%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgca 24
           |||| | |||| |||| |||| |
Db     121 CATTACTATCCACGTCGGGTGTA 99


RESULT  45
US-07-938-782A-1
; Sequence 1, Application US/07938782A
; Patent No. 5525513
;  GENERAL INFORMATION:
;    APPLICANT:  Chen, Jane J.
;    APPLICANT:  London, Irving M.
;    TITLE OF INVENTION:  DNA Encoding the Heme-Regulated
;    TITLE OF INVENTION:  Eurkaryotic Initiation Factor 2 alpha kinase
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Patrea L. Pabst
;      STREET:  2800 One Atlantic Center
;      STREET:  1201 West Peachtree Street
;      CITY:  Atlanta
;      STATE:  Georgia
;      COUNTRY:  US
```

; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/938,782A
; FILING DATE: 31-AUG-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: WU101CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)873-8794
; TELEFAX: (404)873-8795
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2729 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Rabbit
; CELL TYPE: Reticulocytes
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 113..2149
; OTHER INFORMATION: /note= "Expression of HRI
; OTHER INFORMATION: mRNA in Human erythroid cells, using as the
; OTHER INFORMATION: probe rabbit HRI cDNA from nucleotides 113
; OTHER INFORMATION: to 2149."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 229..249
; OTHER INFORMATION: /note= "Primer used in the
; OTHER INFORMATION: amplification of human HRI cDNA sequence
; OTHER INFORMATION: using the rabbit HRI cDNA sequence."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 543..560
; OTHER INFORMATION: /note= "Primer used in the
; OTHER INFORMATION: amplification of human HRI cDNA sequence
; OTHER INFORMATION: using the rabbit HRI cDNA sequence."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 448..468
; OTHER INFORMATION: /note= "Primer used in the
; OTHER INFORMATION: amplification of human HRI cDNA sequence
; OTHER INFORMATION: using the rabbit HRI cDNA sequence."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1009..1031
; OTHER INFORMATION: /note= "Primer used in the
; OTHER INFORMATION: amplification of a human HRI cDNA sequence
; OTHER INFORMATION: using the rabbit HRI cDNA sequence."
; PUBLICATION INFORMATION:
; AUTHORS: Chen, Jane J.
; AUTHORS: London, Irving M.
; TITLE: Cloning of the cDNA of the heme-
; TITLE: regulated eukaryotic initiation factor
; TITLE: 2alpha (eIF-2alpha)kinase of rabbit reticulocytes:
; TITLE: Homology to yeast GCN2 protein kinase and human
; TITLE: double-stranded-RNA-dependent
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 88
; PAGES: 7729-7733
; DATE: September-1991

US-07-938-782A-1

Query Match 60.0%; Score 15; DB 1; Length 2729;
Best Local Similarity 78.3%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 atttcgatcctcgtcaggtgcag 25
||||| | |||| | |||| |||
Db 2059 ATTTCAAGCCTCCTGAGGTTCAG 2081

Search completed: December 26, 2001, 12:27:10
Job time: 7387 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:11:59 ; Search time 4612.2 Seconds
(without alignments)
58.246 Million cell updates/sec

Title: US-09-396-196F-8
Perfect score: 25
Sequence: 1 gcatttcgatcctcgtcaggtgcag 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	19.2	76.8	265	10	BE125672	BE125672 DG1_54_E0
	2	19.2	76.8	353	10	AW287736	AW287736 LG1_271_G
	3	19.2	76.8	542	10	BE357782	BE357782 DG1_22_F0
	4	18.2	72.8	319	11	BG946460	BG946460 PM3-KT004
c	5	18.2	72.8	372	11	W00149	W00149 TgESTzy77e0
c	6	18.2	72.8	431	10	AW925187	AW925187 WS1_76_E0
c	7	18.2	72.8	544	13	AZ143198	AZ143198 SP_0028_B
	8	18.2	72.8	613	13	AQ438224	AQ438224 HS_5083_B
c	9	17.8	71.2	538	10	AV439422	AV439422 AV439422
c	10	17.6	70.4	508	13	AQ358636	AQ358636 HS_5030_A
	11	17.6	70.4	574	10	BE405752	BE405752 WHE1214_H
c	12	17.6	70.4	576	10	BE131409	BE131409 L48-1384T
	13	17.6	70.4	1032	11	BF682210	BF682210 602117637
c	14	17.2	68.8	325	10	BE595789	BE595789 PI1_54_A0
c	15	17.2	68.8	343	10	AW604656	AW604656 CM0-CT031
	16	17.2	68.8	349	11	BF524069	BF524069 UI-R-C0-h
	17	17.2	68.8	364	13	AF116319	AF116319 AF116319
c	18	17.2	68.8	654	10	AA946503	AA946503 EST202002
c	19	17.2	68.8	663	11	BI294795	BI294795 UI-R-DK0-
	20	17.2	68.8	719	10	AW940604	AW940604 GH15883.3
	21	17.2	68.8	988	13	CNS05JPQ	AL340487 Tetraodon
	22	17.2	68.8	997	13	CNS05BPT	AL330122 Tetraodon
	23	17.2	68.8	1081	13	CNS05R7O	AL350205 Tetraodon
	24	17.2	68.8	1095	13	CNS04VWF	AL309624 Tetraodon
	25	17	68.0	113	10	AA422533	AA422533 vf14a04.s
	26	17	68.0	300	11	BF386481	BF386481 UI-R-CA1-
c	27	17	68.0	332	10	AI724043	AI724043 RHIZ1_25_
	28	17	68.0	360	13	AZ817869	AZ817869 2M0087N04
c	29	17	68.0	421	10	AW747232	AW747232 WS1_66_E0
	30	17	68.0	432	11	BF387098	BF387098 UI-R-CA1-
c	31	17	68.0	446	11	BG241150	BG241150 OV1_38_H0
c	32	17	68.0	456	10	AI283774	AI283774 qj75e06.x
c	33	17	68.0	463	11	BG240831	BG240831 OV1_38_H0
c	34	17	68.0	471	11	BG047654	BG047654 OV1_26_E0
c	35	17	68.0	473	11	BG048007	BG048007 OV1_26_E0
	36	17	68.0	476	11	BF388049	BF388049 UI-R-CA1-
c	37	17	68.0	489	10	AW923923	AW923923 WS1_30_A0
c	38	17	68.0	497	10	BE355471	BE355471 DG1_114_H
c	39	17	68.0	502	10	AW747172	AW747172 WS1_66_E0
c	40	17	68.0	506	13	AZ475043	AZ475043 1M0292H17
	41	17	68.0	510	11	BF390169	BF390169 UI-R-CA1-
c	42	17	68.0	511	10	BE364076	BE364076 PI1_11_A0
c	43	17	68.0	511	10	BE364141	BE364141 PI1_11_A0
	44	17	68.0	515	13	AQ940100	AQ940100 Sheared D
c	45	17	68.0	518	10	BE355404	BE355404 DG1_114_H
c	46	17	68.0	519	10	BE355326	BE355326 DG1_114_H
	47	17	68.0	529	13	TA367G09Q	AL495254 T. brucei
	48	17	68.0	535	11	BF392895	BF392895 UI-R-CA0-
	49	17	68.0	537	11	BF387587	BF387587 UI-R-CA1-
	50	17	68.0	546	13	AQ940823	AQ940823 Sheared D
c	51	17	68.0	558	10	AW679714	AW679714 WS1_30_A0
	52	17	68.0	559	11	BF079130	BF079130 229668 MA
	53	17	68.0	585	13	AZ624486	AZ624486 1M0463O11
	54	17	68.0	588	13	AQ766779	AQ766779 HS_2181_A
	55	17	68.0	673	13	AZ796814	AZ796814 2M0052P09
	56	17	68.0	732	10	AI055072	AI055072 coau0002P
c	57	17	68.0	885	13	CNS078XB	AL434565 T3 end of
	58	17	68.0	943	13	CNS04JU8	AL293993 Tetraodon
	59	17	68.0	971	13	CNS0236I	AL179091 Tetraodon
c	60	17	68.0	975	11	BG613958	BG613958 602641677
	61	17	68.0	977	11	BG441880	BG441880 GA__Ea001
c	62	17	68.0	1029	13	CNS05AMS	AL328717 Tetraodon
c	63	17	68.0	1074	13	CNS00EM4	AL069365 Drosophil
	64	16.8	67.2	108	11	BE843877	BE843877 RC0-TN007
	65	16.8	67.2	180	13	CNS06CIC	AL392554 T3 end of
c	66	16.8	67.2	241	10	AW647187	AW647187 EST325730
	67	16.8	67.2	328	10	AI202533	AI202533 qs71h07.x
c	68	16.8	67.2	495	11	BI374546	BI374546 RE62167.5
c	69	16.8	67.2	533	11	BI348313	BI348313 dai28c05.
c	70	16.8	67.2	1072	13	CNS01T0D	AL165910 Tetraodon
c	71	16.6	66.4	222	10	AA359291	AA359291 EST68294
	72	16.6	66.4	276	10	AI252759	AI252759 qh91e03.x
	73	16.6	66.4	292	10	BB310257	BB310257 BB310257
c	74	16.6	66.4	301	11	BG193935	BG193935 RST13075
c	75	16.6	66.4	319	10	BE592622	BE592622 WS1_95_H0
c	76	16.6	66.4	341	10	AI475043	AI475043 tm62e02.x
c	77	16.6	66.4	349	11	BF633421	BF633421 NF056D06D
c	78	16.6	66.4	384	10	AA378341	AA378341 EST91025
c	79	16.6	66.4	415	10	BE362441	BE362441 DG1_86_G1
c	80	16.6	66.4	426	11	BG048729	BG048729 OV1_22_C0
	81	16.6	66.4	429	10	AI235781	AI235781 EST232343
c	82	16.6	66.4	434	11	BF588335	BF588335 FM1_40_E0
	83	16.6	66.4	440	11	BI297398	BI297398 UI-R-CV2-
c	84	16.6	66.4	442	11	BG241314	BG241314 OV1_40_H0
c	85	16.6	66.4	444	11	BG052369	BG052369 RHIZ2_12_

	86	16.6	66.4	445	10	AI716394	AI716394 UI-R-Y0-a
c	87	16.6	66.4	449	10	BE316962	BE316962 NF056B03L
	88	16.6	66.4	456	10	AI412328	AI412328 EST240626
c	89	16.6	66.4	457	11	BF543348	BF543348 UI-R-Y0-a
c	90	16.6	66.4	466	11	BG052299	BG052299 RHIZ2_12_
	91	16.6	66.4	469	10	AI409897	AI409897 EST238190
	92	16.6	66.4	472	11	BF407000	BF407000 UI-R-BJ2-
c	93	16.6	66.4	479	11	BG052298	BG052298 RHIZ2_12_
	94	16.6	66.4	491	11	BG380634	BG380634 UI-R-CT0-
	95	16.6	66.4	497	10	AA891891	AA891891 EST195694
c	96	16.6	66.4	497	10	AW585425	AW585425 EST317048
c	97	16.6	66.4	500	11	C96365	C96365 C96365 Marc
	98	16.6	66.4	501	10	AW922718	AW922718 DG1_45_G1
c	99	16.6	66.4	504	11	BI397990	BI397990 NXPV_113_
c	100	16.6	66.4	507	11	BG049023	BG049023 OV1_22_C0

ALIGNMENTS

```
RESULT   1
BE125672
LOCUS       BE125672      265 bp    mRNA            EST       19-JUL-2000
DEFINITION  DG1_54_E04.b1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BE125672
VERSION     BE125672.1  GI:8548262
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 265)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: dark-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 265
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .265
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Dark Grown 1 (DG1)"
                     /note="Organ: 5-day-old dark-grown seedlings; Vector:
                     Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT       54 a     98 c     75 g     38 t
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 10;  Length 265;
  Best Local Similarity   87.5%;  Pred. No. 1.5e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           |||| || |||||||||| |||||
Db      61 CATTGCGCTCCTCGTCAGCTGCAG 84
```

```
RESULT   2
AW287736
LOCUS       AW287736      353 bp    mRNA            EST       19-JUL-2000
DEFINITION  LG1_271_G07.b1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   AW287736
VERSION     AW287736.2  GI:6859727
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
  TITLE     An EST database from Sorghum: light-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     On Jan 6, 2000 this sequence version replaced gi:6677580.
            Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 352
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .353
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Light Grown 1 (LG1)"
                     /note="Organ: 10- to 14-day-old light-grown (greenhouse)
                     seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
                     ; The library was made from poly-A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT       55 a    120 c    123 g     55 t
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 10;  Length 353;
  Best Local Similarity   87.5%;  Pred. No. 1.6e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           |||| || |||||||||| |||||
Db      47 CATTGCGCTCCTCGTCAGCTGCAG 70
```

```
RESULT   3
BE357782
LOCUS       BE357782      542 bp    mRNA            EST       20-JUL-2000
DEFINITION  DG1_22_F01.b1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BE357782
VERSION     BE357782.1  GI:9299339
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 542)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: dark-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
```

```
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 537
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .542
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Dark Grown 1 (DG1)"
                     /note="Organ: 5-day-old dark-grown seedlings; Vector:
                     Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT       80 a    196 c    187 g     79 t
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 10;  Length 542;
  Best Local Similarity   87.5%;  Pred. No. 1.7e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgcag 25
          |||| || |||||||||| |||||
Db     61 CATTGCGCTCCTCGTCAGCTGCAG 84


RESULT   4
BG946460
LOCUS       BG946460       319 bp    mRNA            EST       11-JUN-2001
DEFINITION  PM3-KT0044-120101-007-d02 KT0044 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG946460
VERSION     BG946460.1  GI:14345831
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 319)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&t2=PM3-KT0044-
            120101-007-d02&t3=2001-01-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 49
            High quality sequence stop: 109.
FEATURES             Location/Qualifiers
     source          1. .319
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="KT0044"
                     /dev_stage="Adult"
                     /note="Organ: bladder_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       74 a     96 c     81 g     68 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 11;  Length 319;
  Best Local Similarity   87.0%;  Pred. No. 4.2e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgca 24
          ||||  |||||||| ||||||||
Db     64 CATTCTGATCCTCGCCAGGTGCA 86


RESULT   5
W00149/c
LOCUS       W00149         372 bp    mRNA            EST       22-MAY-2000
DEFINITION  TgESTzy77e07.r1 TgRH Tachyzoite cDNA Toxoplasma gondii cDNA clone
            tgzy77e07.r1 5', mRNA sequence.
ACCESSION   W00149
VERSION     W00149.1  GI:1271572
KEYWORDS    EST.
SOURCE      Toxoplasma gondii.
  ORGANISM  Toxoplasma gondii
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Sarcocystidae; Toxoplasma.
REFERENCE   1  (bases 1 to 372)
  AUTHORS   Hehl,A., Manger,I., Marra,M., Sibley,L.D., Ajioka,J.A., Aslett,M.A.
            , Dietrich,N., Dubuque,T., Hillier,L., Kucaba,T., Wan,K.L.,
            Waterston,R.H. and Boothroyd,J.
  TITLE     WashU-Merck-Stanford-NIH Toxoplasma EST project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: toxo@watson.wustl.edu
            David Sibley at toxoest@borcim.wustl.edu for further information
            relating to organism, clone or library availability.
            Seq primer: T3
            High quality sequence stop: 321.
FEATURES             Location/Qualifiers
     source          1. .372
                     /organism="Toxoplasma gondii"
                     /strain="RH"
                     /db_xref="taxon:5811"
                     /clone="tgzy77e07.r1"
                     /clone_lib="TgRH Tachyzoite cDNA"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: Lambda ZAP; Site_1: EcoRI; Site_2: XhoI;
                     Toxoplasma RH strain tachyzoites were grown in human
                     foreskin fibroblast cultures in vitro. The library was
                     constructed by K.L. Wan, Cambridge University.  cDNAs were
                     synthesized from polyA RNAs by oligo d(T) priming and
                     directionally cloned into the EcoRI to XhoI sites of the
                     Lambda ZapII vector using the ZAP-cDNA synthesis kit
                     (Stratagene). WARNING: the library contains a small
                     percentage of cDNAs derived from the human host cells."
BASE COUNT      103 a    111 c     85 g     72 t      1 others
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 11;  Length 372;
  Best Local Similarity   87.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atttcgatcctcgtcaggtgcag 25
           |||||| || ||||| |||||||
Db     208 ATTTCGTTCATCGTCTGGTGCAG 186


RESULT   6
AW925187/c
LOCUS       AW925187      431 bp    mRNA            EST       19-JUL-2000
DEFINITION  WS1_76_E02.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW925187
VERSION     AW925187.1  GI:8091013
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 339
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .431
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT       98 a    135 c    106 g     92 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 431;
  Best Local Similarity   87.0%;  Pred. No. 4.4e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgca 24
           ||||  |||||||||||||||| |
Db     425 CATTGGGATCCTCGTCAGGTGGA 403


RESULT   7
AZ143198/c
LOCUS       AZ143198      544 bp    DNA             GSS       28-AUG-2000
DEFINITION  SP_0028_B2_H12_T7 Strongylocentrotus purpuratus, purple sea urchin,
            sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=28 Col=24 Row=P, DNA sequence.
ACCESSION   AZ143198
VERSION     AZ143198.1  GI:8295101
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 28  row: P  column: 24
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 544.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="Plate=28 Col=24 Row=P"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
BASE COUNT      145 a    116 c     87 g    190 t      6 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 13;  Length 544;
  Best Local Similarity   83.3%;  Pred. No. 4.6e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgca 24
           || || |||||||||||  ||||
Db     331 GCCTTGCGATCCTCGTCANATGCA 308


RESULT   8
AQ438224
LOCUS       AQ438224      613 bp    DNA             GSS       31-MAR-1999
DEFINITION  HS_5083_B2_H12_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=659 Col=24 Row=P, DNA sequence.
ACCESSION   AQ438224
VERSION     AQ438224.1  GI:4549563
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
```

```
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 659  row: P  column: 24
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 613.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=659 Col=24 Row=P"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      170 a     97 c    136 g    194 t     16 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 13;  Length 613;
  Best Local Similarity   87.0%;  Pred. No. 4.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atttcgatcctcgtcaggtgcag 25
           ||||| |||||| ||| ||||||
Db     158 ATTTCTATCCTCTTCATGTGCAG 180


RESULT   9
AV439422/c
LOCUS        AV439422      538 bp    mRNA            EST       23-AUG-2000
DEFINITION   AV439422 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
             PS050c10_r 5', mRNA sequence.
ACCESSION    AV439422
VERSION      AV439422.1  GI:8594647
KEYWORDS     EST.
SOURCE       Porphyra yezoensis.
  ORGANISM   Porphyra yezoensis
             Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
             Porphyra.
REFERENCE    1  (bases 1 to 538)
  AUTHORS    Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
             Tabata,S.
  TITLE      Generation of 10,154 expressed sequence tags from a leafy
             gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL    DNA Res. 7, 223-227 (2000)
  MEDLINE    20363100
COMMENT      Contact: Erika Asamizu
             The First Laboratory for Plant Gene Research
             Kazusa DNA Research Institute
             Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
             Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .538
                     /organism="Porphyra yezoensis"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PS050c10_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      114 a    158 c    170 g     96 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 10;  Length 538;
  Best Local Similarity   90.5%;  Pred. No. 6.9e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggt 21
           ||||||| |||| ||||||||
Db     371 GCATTTCCATCCCCGTCAGGT 351


RESULT  10
AQ358636/c
LOCUS        AQ358636      508 bp    DNA             GSS       06-MAR-1999
DEFINITION   HS_5030_A1_C01_T7 RPCI11 Human Male BAC Library Homo sapiens
             genomic clone Plate=606 Col=1 Row=E, DNA sequence.
ACCESSION    AQ358636
VERSION      AQ358636.1  GI:4207512
KEYWORDS     GSS.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 508)
  AUTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
             Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
             Hood,L.
  TITLE      Sequence-tagged connectors: A sequence approach to mapping and
             scanning the human genome
  JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE    99380589
COMMENT      Contact: Mahairas GG, Wallace JC, Hood L
             High Throughput Sequencing Center
             University of Washington
             401 Queen Anne Avenue North, Seattle, WA 98109, USA
             Tel: (206) 616-3618
             Fax: (206) 616-3887
             Email: jwallace@u.washington.edu
             Clones may be purchased from Research Genetics (info@resgen.com).
             BAC end Web Server: http://www.htsc.washington.edu
             Plate: 606  row: E  column: 1
             Seq primer: T7
             Class: BAC ends
             High quality sequence stop: 508.
FEATURES             Location/Qualifiers
     source          1. .508
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=606 Col=1 Row=E"
                     /clone_lib="RPCI11 Human Male BAC Library"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; RPCI11 Human Male BAC Library"
BASE COUNT      147 a     97 c     95 g    153 t     16 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 508;
  Best Local Similarity   80.0%;  Pred. No. 8.4e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||   | |||| ||||||||||||
Db      98 GCANAACTATCCGCGTCAGGTGCAG 74


RESULT  11
BE405752
LOCUS        BE405752      574 bp    mRNA            EST       21-JUL-2000
DEFINITION   WHE1214_H08_O16ZS Wheat etiolated seedling root cDNA library
             Triticum aestivum cDNA clone WHE1214_H08_O16, mRNA sequence.
ACCESSION    BE405752
VERSION      BE405752.1  GI:9365220
KEYWORDS     EST.
SOURCE       bread wheat.
```

```
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Strategene SK primer.
FEATURES             Location/Qualifiers
     source          1. .574
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE1214_H08_O16"
                     /clone_lib="Wheat etiolated seedling root cDNA library"
                     /tissue_type="Root"
                     /dev_stage="Five day old etiolated seedling"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
                     , germinated and grown aseptically in the  dark at room
                     temperature on filter paper with water, nystatin and
                     cefotaxime in covered crystallization dishes. Roots were
                     harvested.  The tissue, total RNA, and poly(A) RNA were
                     prepared, a cDNA library  was made, and the cDNA clones
                     were in vivo excised to give pBluescript  phagemids in the
                     TJ Close lab (Choi, Close, Fenton) at the University  of
                     California, Riverside.  Plasmid DNA preparations and DNA
                     sequencing  were performed in the OD Anderson lab (all
                     other authors)."
BASE COUNT      155 a    125 c    153 g    141 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 574;
  Best Local Similarity   83.3%;  Pred. No. 8.5e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgca 24
           |||||||| | | ||||||| |||
Db     289 GCATTTCGCTGCCCGTCAGGGGCA 312


RESULT  12
BE131409/c
LOCUS       BE131409      576 bp    mRNA            EST       20-FEB-2001
DEFINITION  L48-1384T3 Ice plant Lambda Uni-Zap XR expression library, 48 hours
            NaCl treatment Mesembryanthemum crystallinum cDNA clone L48-1384,
            mRNA sequence.
ACCESSION   BE131409
VERSION     BE131409.1  GI:8578772
KEYWORDS    EST.
SOURCE      common ice plant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Cushman, J.C.
  TITLE     An expressed sequence tag database for the common ice plant,
            Mesembryanthemum crystallinum
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Cushman JC
            Department of Biochemistry
            University of Nevada
            MS200, Reno, NV 89557-0014, USA
            Tel: 775-784-1918
            Fax: 775-784-1650
            Email: jcushman@unr.edu
            PCR PRimers
            FORWARD: T7
            BACKWARD: T3
            Plate: L48-14  row: G  column: 12
            Seq primer: T3
            High quality sequence stop: 350
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
                     /clone="L48-1384"
                     /clone_lib="Ice plant Lambda Uni-Zap XR expression library
                     , 48 hours NaCl treatment"
                     /tissue_type="Leaf, 48 h 0.4M NaCl"
                     /dev_stage="Six week old"
                     /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT      152 a    141 c    118 g    165 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 576;
  Best Local Similarity   83.3%;  Pred. No. 8.5e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgcag 25
           |||||||| | ||||| | |||||
Db     465 CATTTCGACCATCGTCTGCTGCAG 442


RESULT  13
BF682210
LOCUS       BF682210      1032 bp    mRNA            EST       21-DEC-2000
DEFINITION  602117637T1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:3468776 3', mRNA sequence.
ACCESSION   BF682210
VERSION     BF682210.1  GI:11956105
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1032)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: M. Bento Soares, Ph.D.
             cDNA Library Preparation: Bento Soares and M. Fatima Bonaldo
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8476  row: j  column: 09
            High quality sequence start: 21
            High quality sequence stop: 789.
FEATURES             Location/Qualifiers
     source          1. .1032
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3468776"
```

```
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      288 a    261 c    218 g    265 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 1032;
  Best Local Similarity   83.3%;  Pred. No. 9.3e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgcag 25
          |||||| || ||| | ||||||||
Db    810 CATTTCTATTCTCCTTAGGTGCAG 833


RESULT  14
BE595789/c
LOCUS       BE595789      325 bp    mRNA            EST       18-AUG-2000
DEFINITION  PI1_54_A07.b1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BE595789
VERSION     BE595789.1  GI:9850862
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 325)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: pathogen-induced plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 254
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .325
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Pathogen induced 1 (PI1)"
                     /note="Organ: Anthracnose-infected leaves from
                     two-week-old sorghum plants 48 hr after inoculation;
                     Vector: pBluescript II from Lambda Zap II; Site_1: Xho1;
                     Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
                     cultivar) were infected with pathogen (isolate FRM42I of
                     Colletotrichum graminicola, which is a sorghum isolate).
                     RNA was prepared from infected leaves harvested from 45
                     seedlings  48 hours after inoculation.  Note: young
                     seedlings (2 weeks old) exhibit juvenile resistant
                     reaction, which is an incompatible interaction. As they
                     grow older (4 weeks or older), plants resume susceptiblity
                     to anthracnose disease. The library was made from poly-A
                     RNA in the cloning vector lambda ZAP II. Clones to be
                     sequenced were prepared by mass excision. WARNING: While
                     most or all ESTs are expected to derive from the host
                     plant, no effort was made to eliminate ESTs deriving from
                     the pathogen."
BASE COUNT      104 a     62 c     78 g     81 t    .
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 10;  Length 325;
  Best Local Similarity   86.4%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 gcatttcgatcctcgtcaggtg 22
          |||||| ||||||| ||||| |
Db    148 GCATTTTGATCCTCATCAGGAG 127


RESULT  15
AW604656/c
LOCUS       AW604656      343 bp    mRNA            EST       23-MAR-2000
DEFINITION  CM0-CT0312-190100-153-d11 CT0312 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW604656
VERSION     AW604656.1  GI:7309397
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 343)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM0&t2=CM0-CT0312-
            190100-153-d11&t3=2000-01-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 116.
FEATURES             Location/Qualifiers
     source          1. .343
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CT0312"
                     /dev_stage="Adult"
                     /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       77 a    101 c     92 g     73 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 10;  Length 343;
  Best Local Similarity   86.4%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 catttcgatcctcgtcaggtgc 23
          ||||||| || || ||||||||
Db    252 CATTTCGGTCATCTTCAGGTGC 231
```

```
RESULT  16
BF524069
LOCUS       BF524069      349 bp    mRNA            EST       11-DEC-2000
DEFINITION  UI-R-C0-hs-f-07-0-UI.r1 UI-R-C0 Rattus norvegicus cDNA clone
            UI-R-C0-hs-f-07-0-UI 5', mRNA sequence.
ACCESSION   BF524069
VERSION     BF524069.1  GI:11632036
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1792330
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .349
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-C0-hs-f-07-0-UI"
                     /clone_lib="UI-R-C0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C0
                     library is a subtracted library derived from the UI-R-A1
                     and UI-R-E1 libraries. The UI-R-A1 library consisted of a
                     mixture of individually tagged normalized libraries
                     constructed from rat  placenta, adult lung, brain, liver,
                     kidney, heart, spleen, ovary,  and muscle. The UI-R-E1
                     library  consisted of a mixture of individually tagged
                     normalized libraries constructed from 8, 12 and 18-day
                     embryo. The tag is a string of 3-5 nucleotides present
                     between the Not I  site and the oligo-dT track which
                     allows identification of the library of origin of a clone
                     within the mixture. The subtracted library (UI-R-C0) was
                     constructed as follows: PCR amplified cDNA inserts from a
                     pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had
                     been derived was used as a driver in a hybridization with
                     the pooled UI-R-A1 and UI-R-E1 library in the form of
                     single-stranded circles.  The remaining single-stranded
                     circles (subtracted library) was purified by
                     hydroxyapatite column chromatography, converted to
                     double-stranded circles and electroporated into DH10B
                     bacteria (Life Technologies) to generate the UI-R-C0
                     library. This procedure has been previously described
                     (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                     1996)"
BASE COUNT       76 a    105 c     95 g     73 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 11;  Length 349;
  Best Local Similarity   86.4%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgc 23
           || ||| ||||||||||||| ||
Db     295 CACTTCCATCCTCGTCAGGGGC 316


RESULT  17
AF116319
LOCUS       AF116319      364 bp    DNA             GSS       12-APR-2001
DEFINITION  AF116319 Armillaria gallica W40f Armillaria gallica genomic, DNA
            sequence.
ACCESSION   AF116319
VERSION     AF116319.1  GI:4761841
KEYWORDS    GSS.
SOURCE      Armillaria gallica.
  ORGANISM  Armillaria gallica
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
            Agaricales; Tricholomataceae; Armillaria.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   Hodnett,B. and Anderson,J.B.
  TITLE     Genomic stability of two individuals of Armillaria gallica
  JOURNAL   Mycologia 92 (5), 894-899 (2000)
COMMENT     Contact: Anderson JB
            Biology
            University of Toronto at Mississauga
            3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada
            Email: bhodnett@credit.erin.utoronto.ca
            Class: unknown.
FEATURES             Location/Qualifiers
     source          1. .364
                     /organism="Armillaria gallica"
                     /strain="W40f"
                     /db_xref="taxon:47427"
                     /clone_lib="Armillaria gallica W40f"
BASE COUNT       56 a    107 c     71 g    128 t      2 others
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 13;  Length 364;
  Best Local Similarity   86.4%;  Pred. No. 1.2e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgc 23
           ||||||| |||  |||||||||
Db      80 CATTTCGTTCCAAGTCAGGTGC 101


RESULT  18
AA946503/c
LOCUS       AA946503      654 bp    mRNA            EST       08-JAN-1999
DEFINITION  EST202002 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
            ROVAR89 3' end, mRNA sequence.
ACCESSION   AA946503
VERSION     AA946503.1  GI:4132826
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 654)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On May 1, 1998 this sequence version replaced gi:3106419.
            Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
```

```
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .654
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
                     /clone="ROVAR89"
                     /clone_lib="Normalized rat ovary, Bento Soares"
                     /note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoR1;
                     Site_2: NotI"
BASE COUNT      151 a    143 c    198 g    162 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 10;  Length 654;
  Best Local Similarity   86.4%;  Pred. No. 1.3e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgc 23
           || ||| ||||||||||||| ||
Db     575 CACTTCCATCCTCGTCAGGGGC 554


RESULT  19
BI294795/c
LOCUS       BI294795      663 bp    mRNA            EST       19-JUL-2001
DEFINITION  UI-R-DK0-cec-c-03-0-UI.s1 UI-R-DK0 Rattus norvegicus cDNA clone
            UI-R-DK0-cec-c-03-0-UI 3', mRNA sequence.
ACCESSION   BI294795
VERSION     BI294795.1  GI:14957597
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized rat heart pool library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .663
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-DK0-cec-c-03-0-UI"
                     /clone_lib="UI-R-DK0"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DK0
                     library is a subtracted library derived from a mixture of
                     five individually tagged normalized rat libraries:
                     brain-nRBP (20%), heart-nRHP (20%), kidney-nRKP (20%),
                     aorta-nRAP (20%), and placenta-nRPP (20%). Each original
                     library was constructed from a mixture of equal amounts of
                     RNA from seven different developmental time-points:
                     embryonic day 17, embryonic day 19, embryonic day 21,
                     adult day 1, adult day 12, adult day 75, and adult day
                     200. (Exception: the aorta pool does not contain embryonic
                     day 17 RNA and the placenta pool contains only the three
                     embryonic stages). Each library was normalized
                     individually according to the procedure described by
                     Bonaldo, Lennon & Soares (Genome Research Genome 6:
                     791-806, 1996). For construction of the DK0 subtracted
                     library, plasmid DNA from each of the five individually
                     tagged normalized libraries was mixed in the proportions
                     specified above and electroporated into competent bacteria
                     for production of single-stranded circular DNA
                     representing the pool of libraries. Single-stranded
                     circular DNA representing these five normalized libraries
                     was then used as a tracer in a subtractive hybridization
                     with a driver (PCR amplified inserts from a plasmid DNA
                     template preparation) comprising: a) a set of about 1,000
                     arrayed clones from each of the five non-normalized
                     libraries of brain (CT0s), heart (CS0s), kidney (CU0s),
                     aorta (CW0s), and placenta (CX0s). The resulting pool of
                     approximately 5,000 clones represented about 33.3% of the
                     final driver population. A set of about 2,000 arrayed
                     clones from each of the five normalized libraries of brain
                     (CT0), heart (CS0), kidney (CU0), aorta (CW0), and
                     placenta (CX0). The resulting pool of about 10,000 clones
                     represented about 66.6% of the final driver population.
                     TAG_LIB=UI-R-DK0
                     TAG_TISSUE=rat heart pool
                     TAG_SEQ=ATAAGATAAC"
BASE COUNT      147 a    140 c    197 g    179 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 11;  Length 663;
  Best Local Similarity   86.4%;  Pred. No. 1.3e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 catttcgatcctcgtcaggtgc 23
           || ||| ||||||||||||| ||
Db     589 CACTTCCATCCTCGTCAGGGGC 568


RESULT  20
AW940604
LOCUS       AW940604      719 bp    mRNA            EST       23-APR-2001
DEFINITION  GH15883.3prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH15883 3, mRNA sequence.
ACCESSION   AW940604
VERSION     AW940604.1  GI:8116050
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: GH15883.5prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon the presence of a XhoI site followed by a run of 14 or
            more T residues at the beginning of the sequence, this clone was
            polyadenylated. The resulting Poly-T sequence has been removed. hit
            genomic AE003756: Drosophila melanogaster genomic scaffold
```

```
            142000013386035 section 81 of 105, complete sequence.: 02/18/2001
            Plate: GH.158  row: G  column: 11
            High quality sequence stop: 718.
FEATURES              Location/Qualifiers
     source           1. .719
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="GH15883"
                      /clone_lib="GH Drosophila melanogaster head pOT2"
                      /sex="male and female"
                      /dev_stage="adult"
                      /lab_host="DH5 - alpha"
                      /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                      XhoI; Sized fractionated cDNAs were directly ligated into
                      pOT2.  Plasmid cDNA library."
BASE COUNT      162 a    196 c    164 g    197 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 10;  Length 719;
  Best Local Similarity   86.4%;  Pred. No. 1.3e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atttcgatcctcgtcaggtgca 24
           ||| |||| |||| ||||||||
Db     441 ATTGCGATACTCGCCAGGTGCA 462


RESULT  21
CNS05JPQ
LOCUS       CNS05JPQ      988 bp    DNA             GSS       26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            028C23 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL340487
VERSION     AL340487.1  GI:8234245
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 988)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 988)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 988)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES              Location/Qualifiers
     source           1. .988
                      /organism="Tetraodon nigroviridis"
                      /db_xref="taxon:99883"
                      /clone="028C23"
                      /clone_lib="A"
                      /note="Genoscope sequence ID : C0AA028AB12A2-end : T3"
BASE COUNT      281 a    168 c    175 g    331 t     33 others
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 13;  Length 988;
  Best Local Similarity   86.4%;  Pred. No. 1.4e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 tttcgatcctcgtcaggtgcag 25
           ||||| ||||| |||| |||||
Db     796 TTTCGCTCCTCTTCAGTTGCAG 817


RESULT  22
CNS05BPT
LOCUS       CNS05BPT      997 bp    DNA             GSS       26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            002M10 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL330122
VERSION     AL330122.1  GI:8223744
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 997)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 997)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 997)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES              Location/Qualifiers
     source           1. .997
                      /organism="Tetraodon nigroviridis"
                      /db_xref="taxon:99883"
                      /clone="002M10"
                      /clone_lib="A"
                      /note="Genoscope sequence ID : C0AA002BG05C1-end : T7"
BASE COUNT      260 a    222 c    193 g    296 t     26 others
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 13;  Length 997;
  Best Local Similarity   86.4%;  Pred. No. 1.4e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 tttcgatcctcgtcaggtgcag 25
           ||||| ||||| |||| |||||
Db      71 TTTCGCTCCTCTTCAGTTGCAG 92


RESULT  23
CNS05R7O
LOCUS       CNS05R7O     1081 bp    DNA             GSS       26-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
```

```
            043G20 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL350205
VERSION     AL350205.1  GI:8243975
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1081)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1081)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1081)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1081
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="043G20"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA043BD10C1~end : T7"
BASE COUNT      350 a    186 c    188 g    342 t     15 others
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 13;  Length 1081;
  Best Local Similarity   86.4%;  Pred. No. 1.4e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 tttcgatcctcgtcaggtgcag 25
           ||||| ||||| |||| |||||
Db     776 TTTCGCTCCTCTTCAGTTGCAG 797


RESULT  24
CNS04VWF
LOCUS       CNS04VWF     1095 bp    DNA             GSS       24-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            022N08 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL309624
VERSION     AL309624.1  GI:8216563
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1095)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1095)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1095)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1095
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="022N08"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA022DG04A1~end : T3"
BASE COUNT      277 a    244 c    251 g    317 t      6 others
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 13;  Length 1095;
  Best Local Similarity   86.4%;  Pred. No. 1.4e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 tttcgatcctcgtcaggtgcag 25
           ||| |||||| |||||| ||||
Db     596 TTTAGATCCTAGTCAGGAGCAG 617


RESULT  25
AA422533
LOCUS       AA422533      113 bp    mRNA            EST       16-OCT-1997
DEFINITION  vf14a04.s1 Knowles Solter mouse unfertilized egg Mus musculus cDNA
            clone IMAGE:835662 5' similar to SW:MVIN_SALTY P37169 VIRULENCE
            FACTOR MVIN. ;, mRNA sequence.
ACCESSION   AA422533
VERSION     AA422533.1  GI:2101353
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 113)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:495878
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .113
                     /organism="Mus musculus"
```

```
                     /strain="B6D2 F1/J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:835662"
                     /clone_lib="Knowles Solter mouse unfertilized egg"
                     /tissue_type="unfertilized egg"
                     /lab_host="DH10B"
                     /note="Organ: unfertilized egg; Vector: pBluescribe
                     (modified); Site_1: MluI; Site_2: SalI; Cloned
                     unidirectionally from mRNA prepared from 5000 unfertilized
                     eggs.  Primer: SalI(dT):
                     5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTT-3'.  cDNAs were
                     cloned into the MluI/SalI sites of a modified pBluescribe
                     vector using commercial linkers (NEB).  Average insert
                     size:  1.0 kb."
BASE COUNT       15 a      32 c      36 g      30 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 10;  Length 113;
  Best Local Similarity   80.0%;  Pred. No. 1.2e+03;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           ||||  ||||| || |||||||| |
Db       9 GCATGCCGATCTTCTTCAGGTGCGG 33


RESULT  26
BF386481
LOCUS       BF386481      300 bp    mRNA            EST       27-NOV-2000
DEFINITION  UI-R-CA1-bbe-e-03-0-UI.s1 UI-R-CA1 Rattus norvegicus cDNA clone
            UI-R-CA1-bbe-e-03-0-UI 3', mRNA sequence.
ACCESSION   BF386481
VERSION     BF386481.1  GI:11371346
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CA1-bbe-e-03-0-UI"
                     /clone_lib="UI-R-CA1"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA1
                     library is a subtracted library derived from the following
                     tissues: thalamus, cerebellum, hypothalamus, medulla, pons
                     , midbrain, cerebral cortex,  corpus striatum, testis, and
                     hippocampus.  For a detailed description of the  library
                     from which this clone was derived, please visit our web
                     site at  ratest.eng.uiowa.edu.  The subtraction has been
                     previously described in (Bonaldo,  Lennon and Soares,
                     Genome Research 6:791-806, 1996)
                     TAG_SEQ=None found"
BASE COUNT       56 a      81 c      77 g      86 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 11;  Length 300;
  Best Local Similarity   80.0%;  Pred. No. 1.4e+03;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           ||  ||| ||||| ||||||| |||
Db      27 GCTCTTCTATCCTTGTCAGGTACAG 51


RESULT  27
AI724043/c
LOCUS       AI724043      332 bp    mRNA            EST       19-JUL-2000
DEFINITION  RHIZ1_25_E10.y1_A001 Rhizome1 (RHIZ1) Sorghum halepense cDNA, mRNA
            sequence.
ACCESSION   AI724043
VERSION     AI724043.1  GI:5042895
KEYWORDS    EST.
SOURCE      Sorghum halepense.
  ORGANISM  Sorghum halepense
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 332)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Pratt,L.H. and Paterson,A.
  TITLE     An EST database from Sorghum: Sorghum halepense rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence start: 4
            High quality sequence stop: 332
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .332
                     /organism="Sorghum halepense"
                     /db_xref="taxon:4560"
                     /clone_lib="Rhizome1 (RHIZ1)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT       74 a      93 c      75 g      90 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 10;  Length 332;
  Best Local Similarity   80.0%;  Pred. No. 1.4e+03;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||| ||||||||| || ||| || | |
Db     131 GCATGTCGATCCTCCTCTGGAGCCG 107


RESULT  28
AZ817869
```

```
LOCUS       AZ817869      360 bp    DNA              GSS       20-FEB-2001
DEFINITION  2M0087N04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0087N04 R, DNA sequence.
ACCESSION   AZ817869
VERSION     AZ817869.1  GI:12987777
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0087  row: N  column: 04
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 360.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0087N04"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      102 a     79 c     95 g     83 t      1 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 13;  Length 360;
  Best Local Similarity   80.0%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||| || | |||||| |||| |||
Db     102 GCATATCCAGCCTCGTTAGGTTCAG 126


RESULT  29
```

```
AW747232/c
LOCUS       AW747232      421 bp    mRNA             EST       19-JUL-2000
DEFINITION  WS1_66_E04.g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW747232
VERSION     AW747232.1  GI:7660970
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: T7
            High quality sequence stop: 409
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .421
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT       96 a    120 c    111 g     94 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 10;  Length 421;
  Best Local Similarity   80.0%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||| |||||||||| || || || |
Db     244 GCATGTCGATCCTCCTCTGGAGCCG 220


RESULT  30
BF387098
LOCUS       BF387098      432 bp    mRNA             EST       27-NOV-2000
DEFINITION  UI-R-CA1-bbk-c-10-0-UI.s1 UI-R-CA1 Rattus norvegicus cDNA clone
            UI-R-CA1-bbk-c-10-0-UI 3', mRNA sequence.
ACCESSION   BF387098
VERSION     BF387098.1  GI:11371922
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 432)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
```

Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=No.

FEATURES Location/Qualifiers
source 1. .432
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CA1-bbk-c-10-0-UI"
/clone_lib="UI-R-CA1"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA1 library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons , midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_SEQ=None found"

BASE COUNT 76 a 107 c 111 g 138 t
ORIGIN

Query Match 68.0%; Score 17; DB 11; Length 432;
Best Local Similarity 80.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1 gcatttcgatcctcgtcaggtgcag 25
           ||  ||| ||||| ||||||| |||
Db      27 GCTCTTCTATCCTTGTCAGGTACAG 51
```

RESULT 31
BG241150/c
LOCUS BG241150 446 bp mRNA EST 15-FEB-2001
DEFINITION OV1_38_H01.b1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BG241150
VERSION BG241150.1 GI:12776223
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 446)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt ,L.H.
TITLE An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 413
POLYA=No.

FEATURES Location/Qualifiers
source 1. .446
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Ovary 1 (OV1)"
/note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 100 a 128 c 121 g 97 t
ORIGIN

Query Match 68.0%; Score 17; DB 11; Length 446;
Best Local Similarity 80.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||| ||||||||| || || || |
Db     300 GCATGTCGATCCTCCTCTGGAGCCG 276
```

RESULT 32
AI283774/c
LOCUS AI283774 456 bp mRNA EST 28-JAN-1999
DEFINITION qj75e06.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1865314 3' similar to gb:M59979 PROSTAGLANDIN G/H SYNTHASE 1 PRECURSOR (HUMAN);;, mRNA sequence.
ACCESSION AI283774
VERSION AI283774.1 GI:3922007
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 456)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 2267 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .456
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1865314"
/clone_lib="NCI_CGAP_Kid3"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and

M. Fatima Bonaldo. "
BASE COUNT 124 a 128 c 110 g 94 t
ORIGIN

Query Match 68.0%; Score 17; DB 10; Length 456;
Best Local Similarity 80.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcatttcgatcctcgtcaggtgcag 25
||||.|.||||.|.||||.||||||
Db 260 GCATCTGGATCATTGTCATGTGCAG 236

RESULT 33
BG240831/c
LOCUS BG240831 463 bp mRNA EST 15-FEB-2001
DEFINITION OV1_38_H01.g1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BG240831
VERSION BG240831.1 GI:12775904
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 463)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.
TITLE An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: PolyTMix
High quality sequence start: 2
High quality sequence stop: 461
POLYA=No.
FEATURES Location/Qualifiers
source 1. .463
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Ovary 1 (OV1)"
/note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 107 a 131 c 124 g 100 t 1 others
ORIGIN

Query Match 68.0%; Score 17; DB 11; Length 463;
Best Local Similarity 80.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcatttcgatcctcgtcaggtgcag 25
||||.|||||||||.||.||.||.|
Db 300 GCATGTCGATCCTCCTCTGGAGCCG 276

RESULT 34
BG047654/c
LOCUS BG047654 471 bp mRNA EST 25-JAN-2001
DEFINITION OV1_26_E07.b1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BG047654
VERSION BG047654.1 GI:12497673
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 471)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.
TITLE An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 467
POLYA=No.
FEATURES Location/Qualifiers
source 1. .471
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Ovary 1 (OV1)"
/note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 108 a 134 c 125 g 104 t
ORIGIN

Query Match 68.0%; Score 17; DB 11; Length 471;
Best Local Similarity 80.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcatttcgatcctcgtcaggtgcag 25
||||.|||||||||.||.||.||.|
Db 300 GCATGTCGATCCTCCTCTGGAGCCG 276

RESULT 35
BG048007/c
LOCUS BG048007 473 bp mRNA EST 25-JAN-2001
DEFINITION OV1_26_E07.g1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BG048007
VERSION BG048007.1 GI:12498325
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 473)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt,L.H.
TITLE An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860

```
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence stop: 465
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .473
                      /organism="Sorghum bicolor"
                      /db_xref="taxon:4558"
                      /clone_lib="Ovary 1 (OV1)"
                      /note="Organ: Mix of ovaries of varying immature stages
                      from 8-week-old plants; Vector: pBluescript II from Lambda
                      Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                      from poly-A RNA in the cloning vector lambda ZAP II.
                      Clones to be sequenced were prepared by mass excision."
BASE COUNT       109 a    134 c    127 g    103 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 11;  Length 473;
  Best Local Similarity   80.0%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||| ||||||||| || || || |
Db     300 GCATGTCGATCCTCCTCTGGAGCCG 276


RESULT  36
BF388049
LOCUS       BF388049      476 bp    mRNA            EST       27-NOV-2000
DEFINITION  UI-R-CA1-bbx-f-10-0-UI.s1 UI-R-CA1 Rattus norvegicus cDNA clone
            UI-R-CA1-bbx-f-10-0-UI 3', mRNA sequence.
ACCESSION   BF388049
VERSION     BF388049.1  GI:11372873
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .476
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-CA1-bbx-f-10-0-UI"
                      /clone_lib="UI-R-CA1"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA1
                      library is a subtracted library derived from the following
                      tissues: thalamus, cerebellum, hypothalamus, medulla, pons
                      , midbrain, cerebral cortex,  corpus striatum, testis, and
                      hippocampus.  For a detailed description of the  library
                      from which this clone was derived, please visit our web
                      site at  ratest.eng.uiowa.edu.  The subtraction has been
                      previously described in (Bonaldo,  Lennon and Soares,
                      Genome Research 6:791-806, 1996)
                      TAG_SEQ=None found"
BASE COUNT       81 a    117 c    125 g    150 t      3 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 11;  Length 476;
  Best Local Similarity   80.0%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           ||  ||| ||||| ||||||| |||
Db      27 GCTCTTCTATCCTTGTCAGGTACAG 51


RESULT  37
AW923923/c
LOCUS       AW923923      489 bp    mRNA            EST       19-JUL-2000
DEFINITION  WS1_30_A09.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW923923
VERSION     AW923923.1  GI:8089748
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 472
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .489
                      /organism="Sorghum bicolor"
                      /db_xref="taxon:4558"
                      /clone_lib="Water-stressed 1 (WS1)"
                      /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                      water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                      Site_2: EcoRI; The library was made from poly-A RNA in the
                      cloning vector lambda ZAP II.  Clones to be sequenced were
                      prepared by mass excision."
BASE COUNT      120 a    136 c    125 g    108 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 10;  Length 489;
  Best Local Similarity   80.0%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||| ||||||||| || || || |
```

```
Db    343 GCATGTCGATCCTCCTCTGGAGCCG 319


RESULT 38
BE355471/c
LOCUS       BE355471      497 bp    mRNA            EST       20-JUL-2000
DEFINITION  DG1_114_H10.g1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BE355471
VERSION     BE355471.1  GI:9296573
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: dark-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 2
            High quality sequence stop: 475
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Dark Grown 1 (DG1)"
                     /note="Organ: 5-day-old dark-grown seedlings; Vector:
                     Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      125 a    137 c    123 g    111 t      1 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 10;  Length 497;
  Best Local Similarity   80.0%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||| |||||||||| || || || |
Db     335 GCATGTCGATCCTCCTCTGGAGCCG 311


RESULT 39
AW747172/c
LOCUS       AW747172      502 bp    mRNA            EST       19-JUL-2000
DEFINITION  WS1_66_E04.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   AW747172
VERSION     AW747172.1  GI:7660910
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: water-stressed plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 458
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Water-stressed 1 (WS1)"
                     /note="Organ: Mix of 5-week old plants on days 7 & 8 after
                     water was withheld; Vector: Lambda Zap; Site_1: XhoI;
                     Site_2: EcoRI; The library was made from poly-A RNA in the
                     cloning vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      132 a    136 c    130 g    104 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 10;  Length 502;
  Best Local Similarity   80.0%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||| |||||||||| || || || |
Db     304 GCATGTCGATCCTCCTCTGGAGCCG 280


RESULT 40
AZ475043/c
LOCUS       AZ475043      506 bp    DNA             GSS       04-OCT-2000
DEFINITION  1M0292H17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0292H17 R, DNA sequence.
ACCESSION   AZ475043
VERSION     AZ475043.1  GI:10633168
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 506)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0292  row: H  column: 17
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 506.
FEATURES             Location/Qualifiers
     source          1. .506
```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0292H17"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 176 a 106 c 92 g 132 t
ORIGIN

Query Match 68.0%; Score 17; DB 13; Length 506;
Best Local Similarity 80.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcatttcgatcctcgtcaggtgcag 25
|| ||||| ||| || | ||||||||| |||
Db 427 GGATTTTGATACTTGCCAGGTGCAG 403

RESULT 41
BF390169
LOCUS BF390169 510 bp mRNA EST 27-NOV-2000
DEFINITION UI-R-CA1-bcl-c-02-0-U1.s1 UI-R-CA1 Rattus norvegicus cDNA clone UI-R-CA1-bcl-c-02-0-UI 3', mRNA sequence.
ACCESSION BF390169
VERSION BF390169.1 GI:11375013
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 510)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=No.
FEATURES Location/Qualifiers
source 1. .510
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CA1-bcl-c-02-0-UI"
/clone_lib="UI-R-CA1"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA1 library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons , midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_SEQ=None found"
BASE COUNT 90 a 130 c 131 g 158 t 1 others
ORIGIN

Query Match 68.0%; Score 17; DB 11; Length 510;
Best Local Similarity 80.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcatttcgatcctcgtcaggtgcag 25
|| ||| ||||| ||||||| |||
Db 27 GCTCTTCTATCCTTGTCAGGTACAG 51

RESULT 42
BE364076/c
LOCUS BE364076 511 bp mRNA EST 20-JUL-2000
DEFINITION PI1_11_A09.b1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BE364076
VERSION BE364076.1 GI:9305633
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 511)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt ,L.H.
TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 503
POLYA=No.
FEATURES Location/Qualifiers
source 1. .511
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Pathogen induced 1 (PI1)"
/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; Two-week-old sorghum plants (BTX 623 cultivar) were infected with pathogen (isolate FRM42I of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45

seedlings 48 hours after inoculation. Note: young seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptiblity to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."
BASE COUNT 108 a 138 c 116 g 147 t 2 others
ORIGIN

Query Match 68.0%; Score 17; DB 10; Length 511;
Best Local Similarity 80.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcatttcgatcctcgtcaggtgcag 25
|||| ||||||||| || || || |
Db 191 GCATGTCGATCCTCCTCTGGAGCCG 167

RESULT 43
BE364141/c
LOCUS BE364141 511 bp mRNA EST 20-JUL-2000
DEFINITION PI1_11_A09.g1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BE364141
VERSION BE364141.1 GI:9305698
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 511)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt,L.H.
TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: T7
High quality sequence stop: 511
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .511
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Pathogen induced 1 (PI1)"
/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; Two-week-old sorghum plants (BTX 623 cultivar) were infected with pathogen (isolate FRM42I of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings 48 hours after inoculation. Note: young seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptiblity to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."
BASE COUNT 108 a 138 c 116 g 149 t
ORIGIN

Query Match 68.0%; Score 17; DB 10; Length 511;
Best Local Similarity 80.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcatttcgatcctcgtcaggtgcag 25
|||| ||||||||| || || || |
Db 191 GCATGTCGATCCTCCTCTGGAGCCG 167

RESULT 44
AQ940100
LOCUS AQ940100 515 bp DNA GSS 27-JAN-2000
DEFINITION Sheared DNA-43C24.TR Sheared DNA Trypanosoma brucei genomic clone Sheared DNA-43C24, DNA sequence.
ACCESSION AQ940100
VERSION AQ940100.1 GI:6763365
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
REFERENCE 1 (bases 1 to 515)
AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C., Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M.
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: Sheared DNA-43C24.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: http://www.tigr.org/tdb/mdb/tbdb/.
Seq primer: M13-Reverse
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .515
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-43C24"
/clone_lib="Sheared DNA"
/note="Vector: pUC18; Site_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."
BASE COUNT 106 a 158 c 119 g 132 t
ORIGIN

Query Match 68.0%; Score 17; DB 13; Length 515;
Best Local Similarity 80.0%; Pred. No. 1.5e+03;

```
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           || ||||||||| | ||  ||||||
Db      99 GCCTTTCGATCCGCCTCCCGTGCAG 123


RESULT  45
BE355404/c
LOCUS       BE355404      518 bp    mRNA            EST       20-JUL-2000
DEFINITION  DG1_114_H10.b2_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BE355404
VERSION     BE355404.1  GI:9296506
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 518)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: dark-grown seedlings
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 511
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .518
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Dark Grown 1 (DG1)"
                     /note="Organ: 5-day-old dark-grown seedlings; Vector:
                     Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      130 a    149 c    126 g    113 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 10;  Length 518;
  Best Local Similarity   80.0%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gcatttcgatcctcgtcaggtgcag 25
           |||| |||||||||| || || || |
Db     351 GCATGTCGATCCTCCTCTGGAGCCG 327


Search completed: December 26, 2001, 12:12:09
Job time: 6491 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:58:38 ; Search time 1755.64 Seconds
(without alignments)
234.917 Million cell updates/sec

Title: US-09-396-196F-9
Perfect score: 25
Sequence: 1 cgatcctcgtcaggtgcaggtcagc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result No.		Score	Query Match	Length	DB	ID	Description
1		25	100.0	1041	6	AR029499	AR029499 Sequence
2		25	100.0	1041	6	AR034916	AR034916 Sequence
3		25	100.0	1084	6	A11530	A11530 BioB gene o
4		25	100.0	1121	6	E00893	E00893 Genomic DNA
5		25	100.0	5793	1	ECOBIO	J04423 E.coli 7,8-
6		25	100.0	5872	6	A38246	A38246 Sequence 1
7		25	100.0	5872	6	A38251	A38251 Sequence 6
8		25	100.0	5872	6	A93674	A93674 Sequence 1
9		25	100.0	5872	6	A93679	A93679 Sequence 6
10		25	100.0	5872	6	AR101809	AR101809 Sequence
11		25	100.0	5872	6	AR101810	AR101810 Sequence
12		25	100.0	11022	1	AE000180	AE000180 Escherich
13		25	100.0	13501	1	AE005258	AE005258 Escherich
14		25	100.0	297816	1	AP002553	AP002553 Escherich
15		23.4	93.6	5526	1	AF250776	AF250776 Unculture
16		21.8	87.2	7215	1	SMABIO	D17468 S. marcesce
17		20.8	83.2	4000	1	AF250768	AF250768 Unculture
18		20.2	80.8	1041	1	EHBIOTOP2	U38648 Erwinia her
19	c	18.2	72.8	160	9	HSY15922	Y15922 Homo sapien
20		18.2	72.8	525	4	CEU10160	U10160 Cervus elap
21		18.2	72.8	987	6	AR008950	AR008950 Sequence
22		18.2	72.8	987	6	AX154633	AX154633 Sequence
23		18.2	72.8	987	6	I28591	I28591 Sequence 1
24		18.2	72.8	993	4	CEU57752	U57752 Cervus elap
25		18.2	72.8	1005	9	CTU19834	U19834 Cercocebus
26		18.2	72.8	1007	9	AF180563	AF180563 Homo sapi
27		18.2	72.8	1012	4	AF007576	AF007576 Capra hir
28		18.2	72.8	1012	4	BTU11815	U11815 Bos taurus
29		18.2	72.8	1018	6	AR052857	AR052857 Sequence
30		18.2	72.8	1058	4	ECIL12P40	Y11129 Equus cabal
31		18.2	72.8	1080	9	MMU19841	U19841 Macaca mula
32		18.2	72.8	1389	6	AX084106	AX084106 Sequence
33		18.2	72.8	1399	6	AR091391	AR091391 Sequence
34		18.2	72.8	1399	9	HUMCLMF40	M65272 Human cytot
35		18.2	72.8	1560	6	AR091395	AR091395 Sequence
36		18.2	72.8	1623	6	AR091394	AR091394 Sequence
37		18.2	72.8	1870	6	A92079	A92079 Sequence 3
38		18.2	72.8	2318	9	HUMNKSFP40	M65290 Human natur
39		18.2	72.8	2362	6	AR082684	AR082684 Sequence
40		18.2	72.8	2362	6	AR122960	AR122960 Sequence
41		18.2	72.8	2362	6	I28325	I28325 Sequence 1
42		18.2	72.8	2362	6	I89770	I89770 Sequence 1
43	c	18.2	72.8	2611	9	HSAJ2967	AJ222967 Homo sapi
44		18.2	72.8	6139	6	AR091393	AR091393 Sequence
45		18.2	72.8	14698	1	AE007146	AE007146 Mycobacte
46		18.2	72.8	18340	9	AY008847	AY008847 Homo sapi
47		18.2	72.8	33692	2	AC010294	AC010294 Homo sapi
48		18.2	72.8	42729	1	MTCY71	Z92771 Mycobacteri
49		18.2	72.8	91326	2	AC011382	AC011382 Homo sapi
50	c	18.2	72.8	107533	2	AC011418	AC011418 Homo sapi
51		18.2	72.8	159397	2	AC027040	AC027040 Homo sapi
52		18.2	72.8	163594	2	AC027796	AC027796 Homo sapi
53	c	18.2	72.8	200237	9	AF168787	AF168787 Homo sapi
54		17.8	71.2	316	6	AX098150	AX098150 Sequence
55		17.8	71.2	144483	2	AC091701	AC091701 Trypanoso
56		17.8	71.2	155892	2	AP002513	AP002513 Homo sapi
57		17.8	71.2	196613	2	AP002490	AP002490 Homo sapi
58		17.6	70.4	1745	10	AB041601	AB041601 Mus muscu
59	c	17.6	70.4	3929	1	PAOPDEG	Z14064 P.aeruginos
60	c	17.6	70.4	7318	10	AB026497	AB026497 Mus muscu
61		17.6	70.4	11031	1	AE004648	AE004648 Pseudomon
62		17.6	70.4	20445	8	AF034411	AF034411 Lycopersi
63		17.6	70.4	41392	3	AF043706	AF043706 Caenorhab
64	c	17.6	70.4	54841	2	AC017535	AC017535 Drosophil
65	c	17.6	70.4	65036	2	AC015581	AC015581 Homo sapi
66	c	17.6	70.4	69900	2	AC021767	AC021767 Mus muscu
67	c	17.6	70.4	110105	2	AC079502	AC079502 Mus muscu
68	c	17.6	70.4	115382	2	AC010689	AC010689 Drosophil
69		17.6	70.4	116256	9	AL139101	AL139101 Human DNA
70		17.6	70.4	152402	2	AC022658	AC022658 Homo sapi

	71	17.6	70.4	153694	2	AL355359 AL355359 Homo sapi
c	72	17.6	70.4	154804	2	AC022388 AC022388 Homo sapi
c	73	17.6	70.4	165484	2	AL590671 AL590671 Homo sapi
	74	17.6	70.4	168177	2	AC023914 AC023914 Homo sapi
c	75	17.6	70.4	169226	2	AC006913 AC006913 Caenorhab
c	76	17.6	70.4	171723	2	AC020171 AC020171 Drosophil
c	77	17.6	70.4	173844	3	AC008299 AC008299 Drosophil
c	78	17.6	70.4	176249	2	AC069497 AC069497 Homo sapi
c	79	17.6	70.4	195405	2	AC093355 AC093355 Mus muscu
	80	17.6	70.4	196514	9	CNS05TBP AL163952 Human chr
	81	17.6	70.4	197456	2	AC005054 AC005054 Homo sapi
c	82	17.6	70.4	220166	2	AC022781 AC022781 Mus muscu
	83	17.6	70.4	222212	2	AL591065 AL591065 Mus muscu
	84	17.6	70.4	234169	10	MMU400878 AJ400878 Mus muscu
c	85	17.6	70.4	238986	3	AE003774 AE003774 Drosophil
	86	17.6	70.4	295566	3	AE003597 AE003597 Drosophil
c	87	17.2	68.8	4886	9	AB037816 AB037816 Homo sapi
c	88	17.2	68.8	6372	6	AX173175 AX173175 Sequence
c	89	17.2	68.8	8874	9	HSREP10 X00734 Human beta-
	90	17.2	68.8	49779	2	AC083995 AC083995 Homo sapi
	91	17.2	68.8	60231	2	AC090600 AC090600 Homo sapi
	92	17.2	68.8	86896	1	RCU57682 U57682 Rhodobacter
c	93	17.2	68.8	92893	9	AC020610 AC020610 Homo sapi
	94	17.2	68.8	117589	2	AC074267 AC074267 Homo sapi
c	95	17.2	68.8	140791	8	AP002912 AP002912 Oryza sat
c	96	17.2	68.8	141387	9	AC005686 AC005686 Homo sapi
	97	17.2	68.8	141869	2	AC010503 AC010503 Homo sapi
	98	17.2	68.8	149259	9	AC009262 AC009262 Human Chr
	99	17.2	68.8	152195	2	AC016331 AC016331 Homo sapi
c	100	17.2	68.8	156623	2	AL356693 AL356693 Homo sapi

ALIGNMENTS

```
RESULT   1
AR029499
LOCUS       AR029499      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1  GI:5941472
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.Andrew.
  TITLE     Enhanced biotin biosynthesis in plant tissue
  JOURNAL   Patent: US 5859335-A 7 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.25;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    105 CGATCCTCGTCAGGTGCAGGTCAGC 129


RESULT   2
AR034916
LOCUS       AR034916      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1  GI:5950521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.A.
  TITLE     Transgenic plants having increased biotin content
  JOURNAL   Patent: US 5869719-A 7 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.25;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    105 CGATCCTCGTCAGGTGCAGGTCAGC 129


RESULT   3
A11530
LOCUS       A11530       1084 bp    DNA             PAT       09-FEB-1994
DEFINITION  BioB gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1  GI:490218
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   .
  JOURNAL   Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            24. .1064
                     /gene="BIOB"
     CDS             24. .1064
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490219"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
BASE COUNT      271 a    286 c    318 g    209 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1084;
  Best Local Similarity  100.0%;  Pred. No. 0.25;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    128 CGATCCTCGTCAGGTGCAGGTCAGC 152


RESULT   4
E00893
LOCUS       E00893       1121 bp    DNA             PAT       29-SEP-1997
DEFINITION  Genomic DNA encoding biotin Synthetase.
```

```
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES             Location/Qualifiers
     source          1. .1121
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      289 a    296 c    325 g    211 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 0.25;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db     146 CGATCCTCGTCAGGTGCAGGTCAGC 170


RESULT   5
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES             Location/Qualifiers
     source          1. .5793
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     CDS             complement(98. .574)
                     /note="ORF 1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23513.1"
                     /db_xref="GI:455168"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene            complement(633. .1925)
                     /gene="bioA"
     CDS             complement(633. .1925)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase"
                     /protein_id="AAA23514.1"
                     /db_xref="GI:457106"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                     ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                     SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                     GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                     LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                     QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                     PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            2012. .3052
                     /gene="bioB"
     CDS             2012. .3052
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAA23515.1"
                     /db_xref="GI:145425"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3049. .4203
                     /gene="bioF"
     CDS             3049. .4203
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-amino pelargonic acid synthetase"
                     /protein_id="AAA23516.1"
                     /db_xref="GI:145426"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4190. .4945
                     /gene="bioC"
     CDS             4190. .4945
                     /gene="bioC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23517.1"
                     /db_xref="GI:145427"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
```

```
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4938. .5597
                     /gene="bioD"
     CDS             4938. .5597
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAA23518.1"
                     /db_xref="GI:145428"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                     T"
BASE COUNT     1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match             100.0%;  Score 25;  DB 1;  Length 5793;
  Best Local Similarity   100.0%;  Pred. No. 0.22;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db   2116 CGATCCTCGTCAGGTGCAGGTCAGC 2140


RESULT   6
A38246
LOCUS       A38246       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="B1OB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /number=1
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAA02325.1"
                     /db_xref="GI:2294845"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=3
                     /product="PROTEIN"
                     /protein_id="CAA02326.1"
                     /db_xref="GI:2294846"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:
                     8-AMINO-7-OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /number=5
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAA02327.1"
                     /db_xref="GI:2294847"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=6
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAA02328.1"
                     /db_xref="GI:2294848"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
```

```
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.22;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    221 CGATCCTCGTCAGGTGCAGGTCAGC 245


RESULT   7
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.22;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    221 CGATCCTCGTCAGGTGCAGGTCAGC 245


RESULT   8
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
```

```
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKG1GATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN
```

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db     221 CGATCCTCGTCAGGTGCAGGTCAGC 245
```

```
RESULT   9
A93679
LOCUS       A93679       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN
```

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.22;

```
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    221 CGATCCTCGTCAGGTGCAGGTCAGC 245


RESULT  10
AR101809
LOCUS       AR101809     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 1 from patent US 6083712.
ACCESSION   AR101809
VERSION     AR101809.1  GI:12812607
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.22;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    221 CGATCCTCGTCAGGTGCAGGTCAGC 245


RESULT  11
AR101810
LOCUS       AR101810     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 6 from patent US 6083712.
ACCESSION   AR101810
VERSION     AR101810.1  GI:12812608
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.22;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    221 CGATCCTCGTCAGGTGCAGGTCAGC 245


RESULT  12
AE000180
LOCUS       AE000180    11022 bp    DNA             BCT       01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 70 of 400 of the complete
            genome.
ACCESSION   AE000180 U00096
VERSION     AE000180.1  GI:1786988
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 11022)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 11022)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu].  Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data.  Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .11022
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   <1. .130
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 4 REP sequences"
     gene            complement(147. .1430)
                     /gene="ybhC"
                     /note="b0772"
```

```
CDS             complement(147. .1430)
                /gene="ybhC"
                /function="orf; Not classified"
                /note="f427; 98 pct identical to fragment YBHC_ECOLI
                SW:P46130 (300 aa) but contains 127 additional C-terminal
                residues"
                /codon_start=1
                /transl_table=11
                /product="putative pectinesterase"
                /protein_id="AAC73859.1"
                /db_xref="GI:1786989"
                /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPAGTQGVTHTTIQAAVD
                AAIIKRTNKRQYIAVMPGEYQGTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
                NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSNIYYGFLAVNSRFN
                AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
promoter        complement(1507. .1535)
                /note="factor Sigma70; predicted +1 start at 806574"
gene            complement(1582. .2058)
                /gene="ybhB"
                /note="b0773"
CDS             complement(1582. .2058)
                /gene="ybhB"
                /function="orf; Unknown"
                /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC73860.1"
                /db_xref="GI:1786990"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
protein_bind    complement(2108. .2124)
                /note="central position to predicted promoter:85.5"
                /bound_moiety="RhaS predicted site"
gene            complement(2117. .3406)
                /gene="bioA"
                /note="b0774"
CDS             complement(2117. .3406)
                /gene="bioA"
                /EC_number="2.6.1.62"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                /codon_start=1
                /transl_table=11
                /product="7,8-diaminopelargonic acid synthetase"
                /protein_id="AAC73861.1"
                /db_xref="GI:1786991"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
promoter        complement(2193. .2221)
                /gene="bioA"
                /note="factor Sigma70; predicted +1 start at 807260"
protein_bind    3411. .3450
                /note="central position to bioB promoter: -20"
                /bound_moiety="BioB documented site"
protein_bind    3411. .3450
                /note="central position to predicted promoter: -20;genetic
                evidence for the site"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to predicted promoter:50"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to bioA promoter:50"
                /bound_moiety="BioB documented site"
promoter        3413. .3441
                /note="factor Sigma70; promoter bioB; documented +1 at
                808525"
promoter        complement(3447. .3473)
                /note="factor Sigma70; promoter bioA; documented +1 at
                808515"
gene            3493. .4533
                /gene="bioB"
                /note="b0775"
CDS             3493. .4533
                /gene="bioB"
                /EC_number="2.8.1.-"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                /codon_start=1
                /transl_table=11
                /product="biotin synthesis, sulfur insertion?"
                /protein_id="AAC73862.1"
                /db_xref="GI:1786992"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            4530. .5684
                /gene="bioF"
                /note="b0776"
CDS             4530. .5684
                /gene="bioF"
                /EC_number="2.3.1.47"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                /codon_start=1
                /transl_table=11
                /product="8-amino-7-oxononanoate synthase"
                /protein_id="AAC73863.1"
                /db_xref="GI:1786993"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
```

```
  Query Match             100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity   100.0%;  Pred. No. 0.21;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db    3597 CGATCCTCGTCAGGTGCAGGTCAGC 3621


RESULT  13
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
```

```
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .13501
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .7576
                     /note="O-island #36; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655; Cryptic prophage
                     CP-933K; includes one copy of the 13 bp direct repeat that
                     flanks the prophage"
     gene            66. .665
                     /gene="lomK"
                     /note="Z0981"
     CDS             66. .665
                     /gene="lomK"
                     /function="putative membrane; Other or unknown (Phage or
                     Prophage Related)"
                     /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                     residues 1 to 199 of 199 from GenPept 118 :
                     gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
                     [prophage P-EibA]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein of prophage
                     CP-933K"
                     /protein_id="AAG55137.1"
                     /db_xref="GI:12513752"
                     /translation="MRKVCAAILSAATCLAVSGVPAWASEHQSTLSAGYLHASTDAPG
                     SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG
                     PSVRVNEWFSAYAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW
                     GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"
     gene            719. .2041
                     /gene="Z0982"
     CDS             719. .2041
                     /gene="Z0982"
                     /function="putative structure; Structural component (Phage
                     or Prophage Related)"
                     /note="Residues 164 to 440 of 440 are 68.79 pct identical
                     to residues 381 to 645 of 645 from GenPept 118 :
                     gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
                     tail fiber protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tail component of prophage CP-933K"
                     /protein_id="AAG55138.1"
                     /db_xref="GI:12513753"
                     /translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE
                     NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQIGILNDFLGAMTEDDAR
                     PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE
                     SARQAAESAASAKKSEEASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA
                     TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE
                     RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA
                     GPQGPKGETGAAGPVGATGPQGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI
                     TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"
     gene            2004. .2312
                     /gene="Z0984"
     CDS             2004. .2312
                     /gene="Z0984"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 102 of 102 are 98.03 pct identical to
                     residues 28 to 129 of 129 from GenPept 118 :
                     gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
                     hypothetical protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55139.1"
                     /db_xref="GI:12513754"
                     /translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR
                     LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK
                     "
     gene            2489. .3469
                     /gene="Z0985"
     CDS             2489. .3469
                     /gene="Z0985"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 13 to 326 of 326 are 55.73 pct identical
                     to residues 18 to 331 of 336 from GenPept 118 :
                     gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
                     predicted by Glimmer [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55140.1"
                     /db_xref="GI:12513755"
                     /translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF
                     QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS
                     KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR
                     EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD
                     MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD
                     AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"
     gene            3530. .4522
                     /gene="Z0986"
     CDS             3530. .4522
                     /gene="Z0986"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55141.1"
                     /db_xref="GI:12513756"
                     /translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
                     SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK
                     VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY
                     VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG
                     WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT
                     EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"
     gene            5350. .6231
                     /gene="Z0989"
     CDS             5350. .6231
                     /gene="Z0989"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55142.1"
                     /db_xref="GI:12513757"
                     /translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV
```

```
                     KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGEISRKIEVMLNASPVSAAM
                     ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                     FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDIIGIRMDKINGESLLNISSLPAQ
                     AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                     SYHGGKQDLISVVLSKI"
     gene            6462. .7160
                     /gene="Z0990"
     CDS             6462. .7160
                     /gene="Z0990"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55143.1"
                     /db_xref="GI:12513758"
                     /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLRF
                     QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                     IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                     SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                     VSLGFQQVRLHPLL"
     gene            complement(7668. .8144)
                     /gene="ybhB"
                     /note="Z0992"
     CDS             complement(7668. .8144)
                     /gene="ybhB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                     residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                     MG1655: B0773"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55144.1"
                     /db_xref="GI:12513759"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASAS1TAMFS"
     gene            complement(8203. .9492)
                     /gene="bioA"
                     /note="Z0993"
     CDS             complement(8203. .9492)

  Query Match             100.0%;  Score 25;  DB 1;  Length 13501;
  Best Local Similarity   100.0%;  Pred. No. 0.2;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db    9683 CGATCCTCGTCAGGTGCAGGTCAGC 9707


RESULT  14
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa;H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .1245
                     /gene="ECs0753"
     CDS             79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
     gene            1245. .2114
                     /gene="ECs0754"
     CDS             1245. .2114
                     /gene="ECs0754"
                     /note="similar to SUCD_ECOLI gi|1786949 percent identity
                     100 in 289 aa (Conserved in E.coli K-12)"
                     /codon_start=1
```

```
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase alpha subunit"
                     /protein_id="BAB34177.1"
                     /db_xref="GI:13360213"
                     /translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
                     GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
                     LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
                     AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
                     AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
                     LADIGEALKTVLK"
gene                 2781. .3686
                     /gene="ECs0755"
CDS                  2781. .3686
                     /gene="ECs0755"
                     /note="probable transcriptional regulator,similar to
                     transcriptional regulators e.g. glycine cleavage system
                     transcription activator (gcv operon activator) -
                     Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
                     identity 31 in 300 aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative transcription regulator"
                     /protein_id="BAB34178.1"
                     /db_xref="GI:13360214"
                     /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
                     EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
                     PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
                     LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
                     NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
                     EKSDSPHLQQFREWLLAKSVVPQEM"
gene                 complement(3720. .4322)
                     /gene="ECs0756"
CDS                  complement(3720. .4322)
                     /gene="ECs0756"
                     /note="probable cob(I)alamin adenosyltransferase,similar
                     to cob(I)alamin adenosyltransferases (corrinoid
                     adenosyltransferases) e.g. [Escherichia coli]
                     gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
                     aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative cob(I)alamin adenosyltransferase"
                     /protein_id="BAB34179.1"
                     /db_xref="GI:13360215"
                     /translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
                     KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
                     QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
                     IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"
gene                 complement(4332. .5984)
                     /gene="ECs0757"
CDS                  complement(4332. .5984)
                     /gene="ECs0757"
                     /note="probable fumarate hydratase,similar to fumarate
                     hydratases e.g. fumarate hydratase class I, aerobic
                     (fumarase) - Escherichia coli
                     gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
                     aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative fumarate hydratase"
                     /protein_id="BAB34180.1"
                     /db_xref="GI:13360216"
                     /translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
                     ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
                     NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
                     TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
                     GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
                     EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
                     EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
                     DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
                     QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSIGGAAALLAQEYVKSLRCLEYPELG
                     MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"
gene                 complement(6092. .7372)
                     /gene="ECs0758"
CDS                  complement(6092. .7372)
                     /gene="ECs0758"
                     /note="probable transport protein,similar to
                     glutamate/aspartate transport proteins (proton glutamate
                     symport proteins) e.g. [Bacillus stearothermophilus]
                     gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
                     aa,also similar to C4-dicarboxylate transpor"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative transport protein"
                     /protein_id="BAB34181.1"
                     /db_xref="GI:13360217"
                     /translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
                     TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
                     NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
                     FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
                     VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
                     GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
                     LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
                     AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"
gene                 complement(7533. .7850)
                     /gene="ECs0759"
CDS                  complement(7533. .7850)
                     /gene="ECs0759"
                     /note="unknown"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="hypothetical protein"
                     /protein_id="BAB34182.1"
```

```
 Query Match             100.0%;  Score 25;  DB 1;  Length 297816;
 Best Local Similarity   100.0%;  Pred. No. 0.16;
 Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db   92415 CGATCCTCGTCAGGTGCAGGTCAGC 92439


RESULT  15
AF250776
LOCUS       AF250776     5526 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
            modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
            biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
            biotin biosynthesis protein BioC (bioC) genes, complete cds; and
            dethiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION   AF250776
VERSION     AF250776.1  GI:12620124
KEYWORDS    .
SOURCE      uncultured bacterium pCosHE2.
  ORGANISM  uncultured bacterium pCosHE2
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
```

```
source          1. .5526
                /organism="uncultured bacterium pCosHE2"
                /db_xref="taxon:143797"
                /clone="pCosHE2"
                /note="unknown organism, cosmid clone derived from
                environmental consortium"
CDS             complement(52. .528)
                /note="ORF1"
                /codon_start=1
                /transl_table=11
                /product="hypothetical 17.1 kDa protein in modC-bioA
                intergenic region"
                /protein_id="AAG60577.1"
                /db_xref="GI:12620125"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKAG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
gene            complement(587. .1876)
                /gene="bioA"
CDS             complement(587. .1876)
                /gene="bioA"
                /note="7,8-diaminopelargonic acid
                synthetase-aminotransferase"
                /codon_start=1
                /transl_table=11
                /product="DAPA-aminotransferase BioA"
                /protein_id="AAG60578.1"
                /db_xref="GI:12620126"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LKCVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
gene            1963. .3003
                /gene="bioB"
CDS             1963. .3003
                /gene="bioB"
                /codon_start=1
                /transl_table=11
                /product="biotin synthase BioB"
                /protein_id="AAG60579.1"
                /db_xref="GI:12620127"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPNERDMPYLEQMVQGVKDLGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            3000. .4154
                /gene="bioF"
CDS             3000. .4154
                /gene="bioF"
                /note="8-amino-7-oxononanoate synthase"
                /codon_start=1
                /transl_table=11
                /product="KAPA synthetase BioF"
                /protein_id="AAG60580.1"
                /db_xref="GI:12620128"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQLVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                RASLAVIRRDEGDARREKLVSLIAHFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
gene            4141. .4896
                /gene="bioC"
CDS             4141. .4896
                /gene="bioC"
                /note="reaction step prior to synthesis of pimeloyl-CoA"
                /codon_start=1
                /transl_table=11
                /product="biotin biosynthesis protein BioC"
                /protein_id="AAG60581.1"
                /db_xref="GI:12620129"
                /translation="MATVNKQAIAAAFGRAAALYEQHADLQRQSADVLLAMLPERKYT
                HVLDAGCGPGWMTRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                TFDLAWSNLAVQWCGNLSTALRELYRVVRSKGVVAFTTLVQGSLPELHQAWQAVDERP
                HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
gene            4883. .>5526
                /gene="bioD"
CDS             4883. .>5526
                /gene="bioD"
                /note="DTB synthetase"
                /codon_start=1
                /transl_table=11
                /product="dethiobiotin synthetase BioD"
                /protein_id="AAG60582.1"
                /db_xref="GI:12620130"
                /translation="MLVSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVAS
                GSEKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESSVMS
                SGLRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAML
                TAQAIQHAGLTLAGWVANDVTPPGKRHAEYITTLTRMIPAPLLGEIPWLAENPE"
BASE COUNT      1274 a   1507 c   1567 g   1178 t
ORIGIN


  Query Match             93.6%;  Score 23.4;  DB 1;  Length 5526;
  Best Local Similarity   96.0%;  Pred. No. 1.2;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          ||||||||||||||| |||||||||
Db   2067 CGATCCTCGTCAGGTCCAGGTCAGC 2091


RESULT  16
SMABIO
LOCUS       SMABIO       7215 bp    DNA             BCT       04-FEB-1999
DEFINITION  S. marcescens biotin operon, complete sequence.
ACCESSION   D17468
VERSION     D17468.1  GI:402530
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-aminopelargonic acid synthetase; DAPA aminotransferase;
            DTB synthetase; KAPA synthetase; biotin operon; biotin synthetase;
            dethiobiotin synthetase.
SOURCE      Serratia marcescens (strain:Sr41) DNA.
  ORGANISM  Serratia marcescens
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Serratia.
REFERENCE   1  (bases 1 to 7215)
  AUTHORS   Sakurai,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1993) to the DDBJ/EMBL/GenBank databases. Naoki
            Sakurai, Tanabe Seiyaku Co., Ltd., Res Lab of Applied Biochemistry;
            2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan
            (E-mail:GFC01101@niftyserve.or.jp, nsakurai@ddbj.nig.ac.jp,
            Tel:048-433-2545, Fax:048-433-2540)
REFERENCE   2  (bases 1 to 7215)
  AUTHORS   Sakurai,N., Imai,Y., Akatsuka,H., Kawai,E., Komatsubara,S. and
            Tosa,T.
  TITLE     Complete nucleotide sequence of biotin operon of Serratia
            marcescens
  JOURNAL   Unpublished (1993)
COMMENT     Submitted (25-Aug-1993) to DDBJ by:
            Naoki Sakurai
            Res. Lab. of Applied Biochemistry
            Tanabe Seiyaku Co., Ltd.
            2-50 Kawagishi-2-chome
            Toda, Saitama 335
            Japan
```

```
          Phone: 048-433-2545
          Email: nsakurai@ddbj.nig.ac.jp
          Fax:   048-433-2540.
FEATURES             Location/Qualifiers
     source          1. .7215
                     /organism="Serratia marcescens"
                     /strain="Sr41"
                     /db_xref="taxon:615"
     gene            complement(1353. .2630)
                     /gene="bioA"
     CDS             complement(1353. .2630)
                     /gene="bioA"
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase
                     (DAPA aminotransferase)"
                     /protein_id="BAA04284.1"
                     /db_xref="GI:402531"
                     /translation="MSVTASDLAFDQRHIWHPYTSMSRPLPCYPIESASGVELQLADG
                     RSLVDGMSSWWAAIHGYNHPRLNQAASQQLEKMSHVMFGGITHPAAISLCRRLVEMTP
                     EALQCVFLADSGSVAVEVSLKMALQYWQARGERRQRILTLRHGYHGDTFGAMSVCDPD
                     NSMHSLYQGYLAPHLFATAPQCRFDEEWREEDIAPFAALLEQHAGEVAAVILEPVVQG
                     AGGMRIYHPTYLKRVRDVVRVAIKLLLIADEIATGFGRTGKLFACEHAQVVPDILCLG
                     KALTGGYMTLSATLTTRHVAETISNGAAGCFMHGPTFMGNPLACAVADASLALLAENR
                     WQAQVSAIETQLKRELLPLGRCRRWPHVRVLGAIGVVEMREPVDVAELQRGFVERGVW
                     IRPFGKLIYLMPPYIIEAEQLSRLTAAVADAAR"
     gene            2716. .3756
                     /gene="bioB"
     CDS             2716. .3756
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="BAA04285.1"
                     /db_xref="GI:402532"
                     /translation="MMADRIHWTVGQAQALFDKPLLELLFEAQTVHRQHFDPRQVQVS
                     TLLSIKTGACPEDCKYCPQSPRYKTGLESERLMQVEQVLESARKAKANGSTRFCMGAA
                     WKNPHERDMPYLQQMVQGVKAMGMETCMTLGTLDGTQAERLAEAGLDYYNHNLDTSPE
                     FYGSIITTRSYQERLDTLDKVRDAGIKVCSGGIVGLGETVRDRAGLLVQLANLPKPPE
                     SVPINMLVKVKGTPLADNDDVDPFDFIRTIAVARIMMPSSYVRLSAGREQMNEQTQAM
                     CFMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTATEHGDNQQQQVLAKQLL
                     NADTAEFYNAAP"
     gene            3756. .4904
                     /gene="bioF"
     CDS             3756. .4904
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-aminopelargonic acid synthetase (KAPA
                     synthetase)"
                     /protein_id="BAA04286.1"
                     /db_xref="GI:402533"
                     /translation="MSWQQRIEQALAERRLNAAYRRRQTTEGGNGRQIRLGDRLYLNF
                     SGNDYLGLSQDARVIAAWQQGARYGVGNGGSGHVTGFSAAHQALEEQLAAWLGYPRAL
                     LFISYAANQAVLAALMQKGDRILADRLSHASLLEAAGAVAGRRAPVPAQSTAGLARIC
                     LAKPCDGQRLAVTEGLFSMDGDGAPLAELHRLTRAAGAWLMVDDAHGIGVRGEQGRGS
                     CWQQGVRPELLVATFGKAFGVSGAAVLCDEATAEYLLQFARHLIYSTAMPPAQACALQ
                     AALARIREGDDLRARLQDNIRRFRQGAAPLALTLTDSDTAIQPLLVGDNQRALDLATR
                     LRECGLWVSAIRPPTVPPGGARLRLLLTAAHQSQDIDRLLEVLNDVSQ"
     gene            4888. .5655
                     /gene="bioC"
     CDS             4888. .5655
                     /gene="bioC"
                     /function="synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="the product of bioC"
                     /protein_id="BAA04287.1"
                     /db_xref="GI:402534"
                     /translation="MTSANDTVNKQAVASAFSRAAGSYDAAAALQRDVGERLLGMGSS
                     HPGEQLLDAGCGTGYFSRMWRERGKRVTALDLAPGMLDVARQRQAAHHYLLGDIEQVP
                     LPDAAMDICFSSLVVQWCSDLPAALAELYRVTRPGGVILFSTLAAGSLQELGDAWQQV
                     DGERHVNAFLPLTQIRTACAAYRHELVTELRTLNYPDVMTLMRSLKGIGATHLHQGRE
                     GGLMSRGRLAALQAAYPCRQGQFPLSYHLAYGVIYRE"
     gene            5648. .6331
                     /gene="bioD"
     CDS             5648. .6331
                     /gene="bioD"
                     /function="conversion from 7,8-diaminopelargonic acid into
                     dethiobiotin"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase(DTB synthetase)"
                     /protein_id="BAA04288.1"
                     /db_xref="GI:912453"
                     /translation="MSKRWFVTGTDTEVGKSVASSALLQAANRAGYPSAGYKPVASGS
                     EMTAEGLRNGDALALQANSGVALDYDEVNPYVFAEPTSPHIVSADEGRPIDAARLSDG
                     LRRLEQRADWVLVEGAGGWFTPLSAEYTFADWVRQEQLPVILVVGIKLGCINHAVLTA
                     QAVQQAGLTLAGWIANDVGGAPGRRHQEYLATLRRMLPRRCWAKSRTCRRPERAPLGQ
                     YLDISLLAQ"
BASE COUNT     1347 a    2225 c    2367 g    1276 t
ORIGIN


  Query Match             87.2%;  Score 21.8;  DB 1;  Length 7215;
  Best Local Similarity   92.0%;  Pred. No. 7;
  Matches   23;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           ||| || ||||||||||||||||||||
Db    2823 CGACCCGCGTCAGGTGCAGGTCAGC 2847


RESULT  17
AF250768
LOCUS       AF250768     4000 bp    DNA              BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosFS1 BioA-like protein (bioA) gene, partial
            cds; BioB-like protein (bioB) and KAPA-synthetase (bioF) genes,
            complete cds; and biotin biosynthesis protein BioC (bioC) gene,
            partial cds.
ACCESSION   AF250768
VERSION     AF250768.1  GI:12620095
KEYWORDS    .
SOURCE      uncultured bacterium pCosFS1.
  ORGANISM  uncultured bacterium pCosFS1
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 4000)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 4000)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .4000
                     /organism="uncultured bacterium pCosFS1"
                     /db_xref="taxon:143795"
                     /clone="pCosFS1"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     gene            complement(<1. .1177)
                     /gene="bioA"
     CDS             complement(<1. .1177)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="BioA-like protein"
                     /protein_id="AAG60558.1"
                     /db_xref="GI:12620098"
                     /translation="MSVTDSDLEFDQRHIWHPYTSMSHPLPCYPVEAASGVELQLADG
```

```
                     RRLIDGMSSWWAAIHGYNHPQLNLAATQQLEKMSHVMFGGITHPAANELCRKVVAMTP
                     EALQCVFLADSGSVAVEVSLKMALQYWQARGERRQRILTLRHGYHGDTFGAMSVCDPD
                     NSMHSLYQGYLAPHLFATAPQCGFDQAWDENDIAPFAALLEQHAGEVAAVILEPVVQG
                     AGGMRIYHPTYLERVRELCDRYQVLLIADEIATGFGRTGKLFACEHAEVVPDILCLGK
                     ALTGGYMTLSATLTTRHVAETISNGAAGCFMHGPTFMGNPLACAVANASLSLLAENRW
                     QQQVAAIETQLKQQLLPLKMQPKVADVRVLGAIGVVEMREPVDMAQLQRGFVERGVWI
                     "
     gene            1265. .2305
                     /gene="bioB"
     CDS             1265. .2305
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="BioB-like protein"
                     /protein_id="AAG60559.1"
                     /db_xref="GI:12620099"
                     /translation="MMADRIHWTVGLAQTLFDKPLLELLFEAQTVHRQHFDPRQVQVS
                     TLLSIKTGACPEDCKYCPQSSRYKTGLESERLMQVEQVLESARKAKANGSTRFCMGAA
                     WKNPHERDMPYLQQMVQGVKAMGMETCMTLGTLDGSQAERLAEAGLDYYNHNLDTSPE
                     FYGNIITTRSYQERLDTLGKVRGAGIKVCSGGIVGLGETVXXRAGLLVQLANLPTPPE
                     SVPINMLVKVKGTPLADNDDVDPFDFIRTIAVARIMMPSSYVRLSAGREQMNEQTQAM
                     CFMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTETEHGDNQQQEVLAARLM
                     QADTAEFYNAAL"
     gene            2305. .3456
                     /gene="bioF"
     CDS             2305. .3456
                     /gene="bioF"
                     /note="7-keto-8-aminopelargonic acid synthetase; similar
                     to BioF"
                     /codon_start=1
                     /transl_table=11
                     /product="KAPA-synthetase"
                     /protein_id="AAG60556.1"
                     /db_xref="GI:12620096"
                     /translation="MSWQQRIEQALEERXQSAAYRHRLVIKGGNGRYLQLNDQRYLNF
                     SGNDYLGLSQNAQVIAAWQQGAQQYGVGSGGSGHVTGFCTAHQALEQQLAQWLGYPRS
                     LLFISGYAANQALLAALMQGEDRILADRLSHASLLEAAAHSPAQLRRFRHNQPQALAD
                     LLAKPCSGQTLAVTEGVFSMDGDSAPLAELYRLTRDAGAWLMVDDAHGIGVQGEQGRG
                     SCWQQGVRPELLVVTFGKAFGLSGAAVLCDESTAEYLLQFARHLIYSTAMPPAQACAL
                     QAALHCIQQGDELRTRLNTNIQRFRQGAAQLSLTLTESLTAIQPLLVGDNQRALDLAQ
                     HLRGQGLWVSAIRPPTVPPGGARLRITLTAAHQPQDIDRLLEVLHDVSV"
     gene            3440. .>4000
                     /gene="bioC"
     CDS             3440. .>4000
                     /gene="bioC"
                     /note="similar to BioC; probably reaction step prior to
                     synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin biosynthesis protein BioC"
                     /protein_id="AAG60557.1"
                     /db_xref="GI:12620097"
                     /translation="MTSAFEQVNKQAVASAFSRAAVGYDAAAVLQRDVGERLLGMGQE
                     HQGEQLLDAGCGTGYFSRRWRELGKQVTALDLAPGMLAFARQQQAADHYLLGDIENVP
                     LGDAAVDISFSSLVVQWCSDLPRALAELYRVTRPGGVILFSTLAEGSLHELGDAWQQV
                     DGERHVNDFLPLAQIEAACAGYRHHLQ"
BASE COUNT      794 a    1164 c    1236 g     802 t      4 others
ORIGIN


  Query Match             83.2%;  Score 20.8;  DB 1;  Length 4000;
  Best Local Similarity   91.7%;  Pred. No. 22;
  Matches    22;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcagc 25
           ||||| ||||||||||||||| |||
Db    1373 GATCCGCGTCAGGTGCAGGTGAGC 1396


RESULT  18
EHBIOTOP2
LOCUS       EHBIOTOP2     1041 bp    DNA             BCT       04-NOV-1996
DEFINITION  Erwinia herbicola biotin synthetase (bioB) gene, complete cds.


ACCESSION   U38648
VERSION     U38648.1  GI:1228112
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Erwinia herbicola strain=Eho10.
  ORGANISM  Pantoea agglomerans
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Pantoea.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Wu,C.H., Chen,H.Y. and Shiuan,D.
  TITLE     Isolation and characterization of the Erwinia herbicola bio operon
            and the sequences of the bioA and bioB genes
  JOURNAL   Gene 174 (2), 251-258 (1996)
  MEDLINE   97045821
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Shiuan,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-1995) David Shiuan. Biology, Natl. Sun Yat-Sen
            University, Kaohsiung, Taiwan, ROC
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Pantoea agglomerans"
                     /strain="Eho10"
                     /db_xref="taxon:549"
     gene            1. .1041
                     /gene="bioB"
     CDS             1. .1041
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAC44534.1"
                     /db_xref="GI:1228114"
                     /translation="MAHHARWTLSQVTELFDKPLLDLLFEAQTIHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLETERLMEVEQVLDSARKAKNAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLDETQAQRLASAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLDKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSFVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDVQLFRKLGLNPQQTDVMTGDNEQQQKLEEQIFN
                     ADTDQFYNAAAL"
BASE COUNT      230 a    322 c    308 g    181 t
ORIGIN


  Query Match             80.8%;  Score 20.2;  DB 1;  Length 1041;
  Best Local Similarity   88.0%;  Pred. No. 46;
  Matches    22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           ||| ||||| ||||| |||||||||
Db     105 CGACCCTCGCCAGGTTCAGGTCAGC 129


RESULT  19
HSY15922/c
LOCUS       HSY15922      160 bp    DNA             PRI       06-APR-1998
DEFINITION  Homo sapiens CTNS gene, exon 1 and 3' flanking intronic region.
ACCESSION   Y15922
VERSION     Y15922.1  GI:3036848
KEYWORDS    CTNS gene; cystinosin.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160)
  AUTHORS   Antignac,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-1997) Antignac C., U423, INSERM, Tour Lavoisier
            Hopital Necker 149 rue de Sevres Paris, 75015, FRANCE
REFERENCE   2  (bases 1 to 160)
  AUTHORS   Town,M., Jean,G., Cherqui,S., Attard,M., Forestier,L.,
            Whitmore,S.A., Callen,D.F., Gribouval,O., Broyer,M., Bates,G.P.,
            van't Hoff,W. and Antignac,C.
```

```
  TITLE     A novel gene encoding an integral membrane protein is mutated in
            nephropathic cystinosis
  JOURNAL   Nat. Genet. 18 (4), 319-324 (1998)
  MEDLINE   98196758
FEATURES             Location/Qualifiers
     source          1. .160
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     exon            1. .110
                     /gene="CTNS"
                     /number=1
                     /usedin=Y15924:CTNSmRNA
     gene            1. .110
                     /gene="CTNS"
     intron          111. .>160
                     /number=1
BASE COUNT       22 a      55 c      57 g      26 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 9;  Length 160;
  Best Local Similarity   87.0%;  Pred. No. 4.8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           || ||||| ||||||| ||||||
Db      67 GAGCCTCGCCAGGTGCCGGTCAG 45


RESULT  20
CEU10160
LOCUS       CEU10160      525 bp    mRNA            MAM       09-JUL-1994
DEFINITION  Cervus elaphus interleukin 12 mRNA, partial cds.
ACCESSION   U10160
VERSION     U10160.1  GI:508867
KEYWORDS    .
SOURCE      red deer.
  ORGANISM  Cervus elaphus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidea;
            Cervidae; Cervinae; Cervus.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   Lockhart,E.A.
  TITLE     Cloning and sequencing of cervine IL-12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 525)
  AUTHORS   Lockhart,E.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-1994) Euan A. Lockhart, Microbiology, Deer
            Research Laboratory, University of Otago, Union St., Dunedin, New
            Zealand
FEATURES             Location/Qualifiers
     source          1. .525
                     /organism="Cervus elaphus"
                     /db_xref="taxon:9860"
     CDS             <1. .>525
                     /standard_name="IL-12"
                     /codon_start=2
                     /product="interleukin 12"
                     /protein_id="AAA19658.1"
                     /db_xref="GI:508868"
                     /translation="AGQYTCHKGGEVLSRSLLLLHKKEDGIWSTDILKDQKEPKAKSF
                     LKCEAKDYSGHFTCWWLTAISTDLKFSVKSSRGSSDPRGVTCGAMSLSTEKVIVDHRE
                     YKKYTVECQEGSACPAAEESLPIEVVVEAVHKLKYENYTSSFFIRDIIKPDPPKNLQL
                     RPLKNSRQVEVSWE"
BASE COUNT      154 a     120 c     142 g     109 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 4;  Length 525;
  Best Local Similarity   87.0%;  Pred. No. 4.3e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| ||||||| ||||||||
Db     495 ATTCTCGGCAGGTGGAGGTCAGC 517


RESULT  21
AR008950
LOCUS       AR008950      987 bp    DNA             PAT       04-DEC-1998
DEFINITION  Sequence 1 from patent US 5756085.
ACCESSION   AR008950
VERSION     AR008950.1  GI:3967755
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 987)
  AUTHORS   Sykes,M. and Wolf,S.F.
  TITLE     Use of interleukin-12 to prevent graft versus host disease
  JOURNAL   Patent: US 5756085-A 1 26-MAY-1998;
FEATURES             Location/Qualifiers
     source          1. .987
                     /organism="unknown"
BASE COUNT      264 a     243 c     259 g     221 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 6;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| ||||||| ||||||||
Db     743 ATTCTCGGCAGGTGGAGGTCAGC 765


RESULT  22
AX154633
LOCUS       AX154633      987 bp    DNA             PAT       22-JUN-2001
DEFINITION  Sequence 37 from Patent WO0140257.
ACCESSION   AX154633
VERSION     AX154633.1  GI:14536197
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 987)
  AUTHORS   Leong,S.R. and Punnonen,J.
  TITLE     Cytokine polypeptides and nucleic acids
  JOURNAL   Patent: WO 0140257-A 37 07-JUN-2001;
            Maxygen, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .987
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      264 a     243 c     259 g     221 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 6;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| ||||||| ||||||||
Db     743 ATTCTCGGCAGGTGGAGGTCAGC 765


RESULT  23
I28591
LOCUS       I28591        987 bp    DNA             PAT       06-FEB-1997
DEFINITION  Sequence 1 from patent US 5573764.
ACCESSION   I28591
```

```
VERSION     I28591.1  GI:1819367
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 987)
  AUTHORS   Sykes,M. and Wolf,S.F.
  TITLE     Use of interleukin-12 to prevent graft versus host disease
  JOURNAL   Patent: US 5573764-A 1 12-NOV-1996;
FEATURES             Location/Qualifiers
     source          1. .987
                     /organism="unknown"
BASE COUNT      264 a    243 c    259 g    221 t
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 6;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 ATTCTCGGCAGGTGGAGGTCAGC 765
```

```
RESULT  24
CEU57752
LOCUS       CEU57752      993 bp    mRNA            MAM       28-OCT-1999
DEFINITION  Cervus elaphus interleukin-12 p40 subunit mRNA, complete cds.
ACCESSION   U57752
VERSION     U57752.1  GI:1373387
KEYWORDS    .
SOURCE      red deer.
  ORGANISM  Cervus elaphus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidea;
            Cervidae; Cervinae; Cervus.
REFERENCE   1  (bases 1 to 993)
  AUTHORS   Lockhart,E., Slobbe,L. and Buchan,G.
  TITLE     The characterisation of a cervine immunoregulatory cytokine,
            interleukin 12
  JOURNAL   DNA Seq. 10 (2), 139-148 (1999)
  MEDLINE   99304506
   PUBMED   10376218
REFERENCE   2  (bases 1 to 993)
  AUTHORS   Lockhart,E.A. and Buchan,G.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-1996) Euan A. Lockhart, Microbiology, University
            of Otago, Cumberland St, Dunedin, New Zealand
FEATURES             Location/Qualifiers
     source          1. .993
                     /organism="Cervus elaphus"
                     /db_xref="taxon:9860"
                     /cell_type="peripheral blood mononuclear cells"
     primer_bind     1. .33
     CDS             1. .984
                     /codon_start=1
                     /product="interleukin-12 p40 subunit"
                     /protein_id="AAB02258.1"
                     /db_xref="GI:1373388"
                     /translation="MHPQQLVVSWFSLVLLTSPIVAIWELEKNVYVVELDWYPDAPGE
                     TVVLRCDTPEEDGITWTSDQSSEVLGSGKTLTVQVKEFGDAGQYTCHKGGEVLSRSLL
                     LLHKKEDGIWSTDILKDQKEPKAKSFLKCEAKDYSGHFTCWWLTAISTDLKFSVKSSR
                     GSSDPRGVTCGAASLSTEKVIVDHREYKKYTVECQEGSACPAAEESLPIEVVVEAVHK
                     LKYENYTSSFFIRDIIKPDPPKNLQLRPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCV
                     QVQGKNKREKKLFMDQTSAKVTCHKDASIRVQARDRYYNSFWSEWASVSCS"
     primer_bind     complement(961. .993)
BASE COUNT      269 a    240 c    263 g    221 t
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 4;  Length 993;
  Best Local Similarity   87.0%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     746 ATTCTCGGCAGGTGGAGGTCAGC 768
```

```
RESULT  25
CTU19834
LOCUS       CTU19834      1005 bp    mRNA            PRI       03-FEB-1996
DEFINITION  Cercocebus torquatus interleukin-12 alpha p40 subunit (IL-12a)
            mRNA, complete cds.
ACCESSION   U19834
VERSION     U19834.1  GI:644785
KEYWORDS    .
SOURCE      red-crowned mangabey.
  ORGANISM  Cercocebus torquatus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Cercocebus.
REFERENCE   1  (bases 1 to 1005)
  AUTHORS   Villinger,F., Brar,S.S., Mayne,A., Chikkala,N. and Ansari,A.A.
  TITLE     Comparative sequence analysis of cytokine genes from human and
            nonhuman primates
  JOURNAL   J. Immunol. 155 (8), 3946-3954 (1995)
  MEDLINE   96003435
REFERENCE   2  (bases 1 to 1005)
  AUTHORS   Villinger,F.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1995) Francois J. Villinger, Emory University,
            Pathology, 1327 Clifton Rd, Atlanta, GA 30322, USA
FEATURES             Location/Qualifiers
     source          1. .1005
                     /organism="Cercocebus torquatus"
                     /sub_species="atys"
                     /db_xref="taxon:9530"
                     /clone="FPn IL-12a"
                     /cell_type="peripheral blood mononuclear cells"
                     /tissue_type="blood"
     gene            9. .995
                     /gene="IL-12a"
     CDS             9. .995
                     /gene="IL-12a"
                     /codon_start=1
                     /function="NK cell stimulatory factor"
                     /product="interleukin-12 alpha p40 subunit"
                     /protein_id="AAA86701.1"
                     /db_xref="GI:644786"
                     /translation="MCHQQLVISWFSLVFLASPLMAIWELKKDVYVVELDWYPDAPGE
                     MVVLTCDTPEEDGITWTLDQSGEVLGSGKTLTIQVKEFGDAGQYTCHKGGEALSHSLL
                     LPHKKEDGIWSTDILKDQKEPKNETFLRCEAKNYSGRITCWWLSTISTDLTFSIISSR
                     GSSNPQGVTCGAATLSAERVRGDNKEYEYSVECQEDSACPAAEERLPIEVMVDAIHKL
                     KYENYTSSFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCIQ
                     VQGKSKREKKDRIFTDKTSATVICRKNASFSVQAQDRYYSSSWNEWTSVPCS"
BASE COUNT      275 a    252 c    257 g    221 t
ORIGIN
```

```
  Query Match             72.8%;  Score 18.2;  DB 9;  Length 1005;
  Best Local Similarity   87.0%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     751 ATTCTCGGCAGGTGGAGGTCAGC 773
```

```
RESULT  26
AF180563
LOCUS       AF180563      1007 bp    mRNA            PRI       26-SEP-1999
DEFINITION  Homo sapiens interleukin 12, P40 mRNA, complete cds.
ACCESSION   AF180563
VERSION     AF180563.1  GI:5923854
```

```
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1007)
  AUTHORS   Hongyuan,J. and Meiyun,Z.
  TITLE     Cloning and sequence analysis of IL-12 cDNA from Chinese
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1007)
  AUTHORS   Hongyuan,J. and Meiyun,Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1999) Hepatitis, Institute of Virology, CAPM, 100
            Yingxinjie, Xuanwuqu, Beijing 100052, China
FEATURES             Location/Qualifiers
     source          1. .1007
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="peripheral mononuclear"
     CDS             7. .993
                     /note="NKSF/CLMF; IL12p40; 40kD"
                     /codon_start=1
                     /product="interleukin 12, P40"
                     /protein_id="AAD56386.1"
                     /db_xref="GI:5923855"
                     /translation="MCHQQLVISWFSLVFLASPLVAIWELKKDVYVVELDWYPDAPGE
                     MVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVLSHSLL
                     LLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWWLTTISTDLTFSVKSSR
                     GSSDPQGVTCGAATLSAERVRGDNKEYEYSVECQEDSACPAAEESLPIEVMVDAVHKL
                     KYENYTSSFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCVQ
                     VQGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPCS"
BASE COUNT      270 a    247 c    265 g    225 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 9;  Length 1007;
  Best Local Similarity   87.0%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     749 ATTCTCGGCAGGTGGAGGTCAGC 771


RESULT  27
AF007576
LOCUS       AF007576     1012 bp    mRNA            MAM       14-JUL-1997
DEFINITION  Capra hircus interleukin 12 p40 subunit (IL-12) mRNA, complete cds.
ACCESSION   AF007576
VERSION     AF007576.1  GI:2253433
KEYWORDS    .
SOURCE      goat.
  ORGANISM  Capra hircus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Caprinae; Capra.
REFERENCE   1  (bases 1 to 1012)
  AUTHORS   Beyer,J.C., Cheevers,W.P. and Kumpula-McWhirter,N.M.
  TITLE     CLoning of caprine IL-12 40 kDa subunit
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1012)
  AUTHORS   Beyer,J.C., Cheevers,W.P. and Kumpula-McWhirter,N.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-1997) Veterinary Microbiology and Pathology,
            Washington State University, P.O. Box 647040, Pullman, WA
            99164-7040, USA
FEATURES             Location/Qualifiers
     source          1. .1012
                     /organism="Capra hircus"
                     /db_xref="taxon:9925"
                     /cell_type="LPS-stimulated microglial cell"
     gene            1. .1012
                     /gene="IL-12"
     CDS             1. .984
                     /gene="IL-12"
                     /codon_start=1
                     /product="interleukin 12 p40 subunit"
                     /protein_id="AAB62945.1"
                     /db_xref="GI:2253434"
                     /translation="MHPQQLVVSWFSLVLLASPIVAIWELEKNVYVVELDWYPNAPGE
                     TVVLTCDTPEEDGITWTSDQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVLSRSLL
                     LLHKKEDGIWSTDILKDQKEPKAKSFLKCEAKDYSGHFTCSWLTAISTNLKFSVKSSR
                     GSSDPRGVTCGAASLSAEKVSMDHREYNKYTVECQEGSACPAAEESLPIEVVMEAVHK
                     LKYENYTSSFFIRDIIKPDPPKNLQLRPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCV
                     QVQGKNKREKKLFTDQTSAKVTCHKDANIRVQARDRYYSSFWSEWASVSCS"
BASE COUNT      280 a    253 c    260 g    219 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 4;  Length 1012;
  Best Local Similarity   87.0%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     746 ATTCTCGGCAGGTGGAGGTCAGC 768


RESULT  28
BTU11815
LOCUS       BTU11815     1012 bp    mRNA            MAM       27-JAN-1996
DEFINITION  Bos taurus interleukin 12 40 kDa subunit mRNA, complete cds.
ACCESSION   U11815
VERSION     U11815.1  GI:555795
KEYWORDS    .
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1012)
  AUTHORS   Zarlenga,D.S., Canals,A., Aschenbrenner,R.A. and Gasbarre,L.C.
  TITLE     Enzymatic amplification and molecular cloning of cDNA encoding the
            small and large subunits of bovine interleukin 12
  JOURNAL   Biochim. Biophys. Acta 1270 (2-3), 215-217 (1995)
  MEDLINE   95244628
REFERENCE   2  (bases 1 to 1012)
  AUTHORS   Zarlenga,D.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-1994) Dante S. Zarlenga, U.S. Department of
            Agriculture, ARS, LPSI, PIL, Beltsville, MD 20705, USA
FEATURES             Location/Qualifiers
     source          1. .1012
                     /organism="Bos taurus"
                     /strain="Holstein"
                     /db_xref="taxon:9913"
                     /clone="BIL12 40"
                     /sex="male"
                     /cell_type="lymphocytes and macrophages"
                     /tissue_type="lymphoid"
     CDS             1. .984
                     /codon_start=1
                     /product="interleukin 12 40 kDa subunit"
                     /protein_id="AAA85792.1"
                     /db_xref="GI:555796"
                     /translation="MHPQQLVVSWFSLVLLASPIVAMWELEKNVYVVELDWYPDAPGE
                     TVVLTCDTPEEDGITWTSDQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEALSRSLL
                     LLHKKEDGIWSTDILKDQKEPKAKSFLKCEAKDYSGHFTCWWLTAISTDLKFSVKSSR
                     GSSDPRGVTCGAALLSAEKVSLEHREYNKYTVECQEGSACPAAEESLLIEVVVEAVHK
                     LKYENYTSSFFIRDIIKPDPPKNLQLRPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCV
                     QVQGKNKREKKLFMDQTSAKVTCHKDANVRVQARDRYYSSFWSEWASVSCS"
BASE COUNT      275 a    245 c    267 g    225 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 4;  Length 1012;
```

```
  Best Local Similarity   87.0%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    746 ATTCTCGGCAGGTGGAGGTCAGC 768


RESULT  29
AR052857
LOCUS       AR052857     1018 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 194 from patent US 5833975.
ACCESSION   AR052857
VERSION     AR052857.1  GI:5977719
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1018)
  AUTHORS   Paoletti,E., Tartaglia,J. and Cox,W.I.
  TITLE     Canarypox virus expressing cytokine and/or tumor-associated antigen
            DNA sequence
  JOURNAL   Patent: US 5833975-A 194 10-NOV-1998;
FEATURES             Location/Qualifiers
     source          1. .1018
                     /organism="unknown"
BASE COUNT      283 a    245 c    260 g    230 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 6;  Length 1018;
  Best Local Similarity   87.0%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    774 ATTCTCGGCAGGTGGAGGTCAGC 796


RESULT  30
ECIL12P40
LOCUS       ECIL12P40    1058 bp    mRNA            MAM       17-NOV-1999
DEFINITION  Equus caballus mRNA for interleukin 12 p40 subunit.
ACCESSION   Y11129
VERSION     Y11129.1  GI:5441622
KEYWORDS    interleukin-12 p40.
SOURCE      horse.
  ORGANISM  Equus caballus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE   1  (bases 1 to 1058)
  AUTHORS   Nicolson,L., Penha-Goncalves,M.N., Keanie,J.L., Logan,N.A.,
            Argyle,D.J. and Onions,D.E.
  TITLE     Cloning and sequencing of horse interleukin-12 and interleukin-18
            cDNAs
  JOURNAL   Immunogenetics 50 (1-2), 94-97 (1999)
  MEDLINE   20009534
REFERENCE   2  (bases 1 to 1058)
  AUTHORS   Nicolson,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-1997) L. Nicolson, Dept. of Vet. Pathology,
            University of Glasgow Vet. School, Bearsden Rd., Glasgow G61 1QH,
            UK
FEATURES             Location/Qualifiers
     source          1. .1058
                     /organism="Equus caballus"
                     /db_xref="taxon:9796"
                     /cell_type="alveolar macrophages"
                     /cell_line="LPS-stimulated alveolar macrophages"
     gene            1. .990
                     /gene="IL-12 p40"
     misc_feature    1. .12
                     /gene="IL-12 p40"
                     /note="primer-derived sequence"
     CDS             1. .990
                     /gene="IL-12 p40"
                     /codon_start=1
                     /product="interleukin 12 p40 subunit"
                     /protein_id="CAA72011.1"
                     /db_xref="GI:5441623"
                     /translation="MCHQWLVLSWFSLVLLASPLMAIWELEKDVYVVELDWYPDAPGE
                     MVVLTCNTPEEEGITWTSAQSNEVLGSGKTLTIQVKEFGDAGWYTCHKGGEVLSHSHL
                     LLHKKEDGIWSTDILKDQKESKNKTFLKCEAKNYSGRFTCWWLTAISTDLKFSVKSSR
                     GSSDPRGVTCGAATLSAERVSVDDREYKKYTVECQEGSACPAAEESLPIEIVVDAVHK
                     LKYENYTSGFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPETWSTPHSYFSLTFSI
                     QVQGKNKKERKDRLFMDETSATVTCHKDGQIRVQARDRYYSSSWSEWASVSCS"
     sig_peptide     1. .66
                     /gene="IL-12 p40"
     variation       714
                     /gene="IL-12 p40"
                     /note="allelic difference, no change in amino acid
                     sequence"
                     /replace="g"
     misc_feature    1041. .1058
                     /note="primer-derived sequence"
BASE COUNT      284 a    269 c    282 g    223 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 4;  Length 1058;
  Best Local Similarity   87.0%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    746 ATTCTCGGCAGGTGGAGGTCAGC 768


RESULT  31
MMU19841
LOCUS       MMU19841     1080 bp    mRNA            PRI       03-FEB-1996
DEFINITION  Macaca mulatta interleukin-12 alpha (IL-12a) mRNA, complete cds.
ACCESSION   U19841
VERSION     U19841.1  GI:644799
KEYWORDS    .
SOURCE      rhesus monkey.
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1080)
  AUTHORS   Villinger,F., Brar,S.S., Mayne,A., Chikkala,N. and Ansari,A.A.
  TITLE     Comparative sequence analysis of cytokine genes from human and
            nonhuman primates
  JOURNAL   J. Immunol. 155 (8), 3946-3954 (1995)
  MEDLINE   96003435
REFERENCE   2  (bases 1 to 1080)
  AUTHORS   Villinger,F.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1995) Francois J. Villinger, Emory University,
            Pathology, 1327 Clifton Rd, Atlanta, GA 30322, USA
FEATURES             Location/Qualifiers
     source          1. .1080
                     /organism="Macaca mulatta"
                     /db_xref="taxon:9544"
                     /clone="RSo IL-12a clone 4"
                     /cell_type="peripheral blood mononuclear cells"
                     /tissue_type="blood"
     gene            13. .999
                     /gene="IL-12a"
     CDS             13. .999
                     /gene="IL-12a"
                     /codon_start=1
                     /function="NK cell stimulatory factor"
                     /product="interleukin-12 alpha"
                     /protein_id="AAB60397.1"
```

```
                     /db_xref="GI:644800"
                     /translation="MCHQQLVISWFSLVFLASPLMAIWELKKDVYVVELDWYPDAPGE
                     MVVLTCDTPEEDGITWTLDQSGEVLGSGKTLTIQVKEFGDAGQYTCHKGGEALSHSLL
                     LLHKKEDGIWSTDVLKDQKEPKNKTFLRCEAKNYSGRFTCWWLTTISTDLTFSVKSSR
                     GSSNPQGVTCGAVTLSAERVRGDNKEYEYSVECQEDSACPAAEERLPIEVMVDAIHKL
                     KYENYTSSFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCIQ
                     VQGKSKREKKDRIFTDKTSATVICRKNASFSVQAQDRYYSSSWSEWASVPCS"
     variation       329
                     /gene="IL-12a"
                     /replace="t"
     variation       405
                     /gene="IL-12a"
                     /replace="c"
     variation       519
                     /gene="IL-12a"
                     /replace="a"
     variation       629
                     /gene="IL-12a"
                     /note="Leu to Pro"
                     /replace="c"
     variation       665
                     /gene="IL-12a"
                     /note="Leu to Pro"
                     /replace="c"
     variation       714
                     /gene="IL-12a"
                     /replace="g"
     variation       878
                     /gene="IL-12a"
                     /note="Ile to Thr"
                     /replace="c"
     variation       895
                     /gene="IL-12a"
                     /note="Ser to Pro"
                     /replace="c"
BASE COUNT      309 a    263 c    276 g    232 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 9;  Length 1080;
  Best Local Similarity   87.0%;  Pred. No. 4.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    755 ATTCTCGGCAGGTGGAGGTCAGC 777


RESULT  32
AX084106
LOCUS       AX084106     1389 bp    DNA             PAT       28-FEB-2001
DEFINITION  Sequence 4 from Patent WO0110912.
ACCESSION   AX084106
VERSION     AX084106.1  GI:13185656
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 1389)
  AUTHORS   Gillies,S.D. and Lo,K.M.
  TITLE     Multiple cytokine-antibody complexes
  JOURNAL   Patent: WO 0110912-A 4 15-FEB-2001;
            Lexigen Pharmaceuticals Corp. (US)
FEATURES             Location/Qualifiers
     source          1. .1389
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="human p40-IL-2 fusion protein coding sequence"
BASE COUNT      413 a    325 c    327 g    324 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 6;  Length 1389;
  Best Local Similarity   87.0%;  Pred. No. 4e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    743 ATTCTCGGCAGGTGGAGGTCAGC 765


RESULT  33
AR091391
LOCUS       AR091391     1399 bp    DNA             PAT       07-SEP-2000
DEFINITION  Sequence 3 from patent US 5994104.
ACCESSION   AR091391
VERSION     AR091391.1  GI:10018146
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1399)
  AUTHORS   Anderson,R.James, Prentice,H.Grant and MacDonald,I.Duncan.
  TITLE     Interleukin-12 fusion protein
  JOURNAL   Patent: US 5994104-A 3 30-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .1399
                     /organism="unknown"
BASE COUNT      390 a    310 c    333 g    366 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 6;  Length 1399;
  Best Local Similarity   87.0%;  Pred. No. 4e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    785 ATTCTCGGCAGGTGGAGGTCAGC 807


RESULT  34
HUMCLMF40
LOCUS       HUMCLMF40    1399 bp    mRNA            PRI       27-APR-1993
DEFINITION  Human cytotoxic lymphocyte maturation factor 40 kDa subunit mRNA,
            complete cds.
ACCESSION   M65272 M38443 M38444
VERSION     M65272.1  GI:180625
KEYWORDS    cytotoxic lymphocyte maturation factor 40 kDa subunit; interleukin.
SOURCE      Human, cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1399)
  AUTHORS   Gubler,U., Chua,A.O., Schoenhaut,D.S., Dwyer,C.M., McComas,W.,
            Motyka,R., Nabavi,N., Wolitzky,A.G., Quinn,P.M., Familletti,P.C.
            and Gately,M.K.
  TITLE     Coexpression of two distinct genes is required to generate secreted
            bioactive cytotoxic lymphocyte maturation factor
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88, 4143-4147 (1991)
  MEDLINE   91239523
FEATURES             Location/Qualifiers
     source          1. .1399
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             43. .1029
                     /codon_start=1
                     /product="cytotoxic lymphocyte maturation factor 40 kDa
                     subunit"
                     /protein_id="AAA35695.1"
                     /db_xref="GI:180626"
                     /translation="MCHQQLVISWFSLVFLASPLVAIWELKKDVYVVELDWYPDAPGE
                     MVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVLSHSLL
                     LLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWWLTTISTDLTFSVKSSR
                     GSSDPQGVTCGAATLSAERVRGDNKEYEYSVECQEDSACPAAEESLPIEVMVDAVHKL
                     KYENYTSSFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCVQ
```

VQGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPCS"
BASE COUNT 390 a 310 c 333 g 366 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 9; Length 1399;
Best Local Similarity 87.0%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 785 ATTCTCGGCAGGTGGAGGTCAGC 807

RESULT 35
AR091395
LOCUS AR091395 1560 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 11 from patent US 5994104.
ACCESSION AR091395
VERSION AR091395.1 GI:10018150
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1560)
AUTHORS Anderson,R.James, Prentice,H.Grant and MacDonald,I.Duncan.
TITLE Interleukin-12 fusion protein
JOURNAL Patent: US 5994104-A 11 30-NOV-1999;
FEATURES Location/Qualifiers
source 1. .1560
/organism="unknown"
BASE COUNT 443 a 375 c 393 g 349 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 6; Length 1560;
Best Local Similarity 87.0%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 680 ATTCTCGGCAGGTGGAGGTCAGC 702

RESULT 36
AR091394
LOCUS AR091394 1623 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 9 from patent US 5994104.
ACCESSION AR091394
VERSION AR091394.1 GI:10018149
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1623)
AUTHORS Anderson,R.James, Prentice,H.Grant and MacDonald,I.Duncan.
TITLE Interleukin-12 fusion protein
JOURNAL Patent: US 5994104-A 9 30-NOV-1999;
FEATURES Location/Qualifiers
source 1. .1623
/organism="unknown"
BASE COUNT 449 a 393 c 408 g 373 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 6; Length 1623;
Best Local Similarity 87.0%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 743 ATTCTCGGCAGGTGGAGGTCAGC 765

RESULT 37
A92079
LOCUS A92079 1870 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 3 from Patent WO9821322.
ACCESSION A92079
VERSION A92079.1 GI:6740883
KEYWORDS .
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1870)
AUTHORS Junghans,C. and Wittig,B.
TITLE DESIGN PRINCIPLE FOR CONSTRUCTING EXPRESSION CONSTRUCTS FOR GENE THERAPY
JOURNAL Patent: WO 9821322-A 3 22-MAY-1998;
JUNGHANS CLAAS (DE); SOFT GENE GMBH (DE)
FEATURES Location/Qualifiers
source 1. .1870
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 500 a 456 c 445 g 469 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 6; Length 1870;
Best Local Similarity 87.0%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 1375 ATTCTCGGCAGGTGGAGGTCAGC 1397

RESULT 38
HUMNKSFP40
LOCUS HUMNKSFP40 2318 bp mRNA PRI 07-JAN-1995
DEFINITION Human natural killer cell stimulatory factor (NKSF) mRNA, complete cds, clone p40.
ACCESSION M65290
VERSION M65290.1 GI:189235
KEYWORDS cytokine; natural killer cell stimulatory factor.
SOURCE Human EBV-transformed B cell line RPMI 8866, cDNA to mRNA, clone p40.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2318)
AUTHORS Wolf,S.F., Temple,P.A., Kobayashi,M., Young,D., Dicig,M., Lowe,L., Dzialo,R., Fitz,L., Ferenz,C., Hewick,R.M., Kelleher,K., Herrmann,S.H., Clark,S.C., Azzoni,L., Chan,S.H., Trinchieri,G. and Perussia,B.
TITLE Cloning of cDNA for natural killer cell stimulatory factor, a heterodimeric cytokine with multiple biologic effects on T and natural killer cells
JOURNAL J. Immunol. 146 (9), 3074-3081 (1991)
MEDLINE 91201875
FEATURES Location/Qualifiers
source 1. .2318
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="EBV-transformed B cell line RPMI 8866"
/map="Unassigned"
sig_peptide 14. .79
/gene="NKSF2"
/note="G00-127-870"
CDS 14. .1000
/gene="NKSF2"
/codon_start=1
/db_xref="GDB:G00-127-870"
/product="natural killer cell stimulatory factor"
/protein_id="AAA59938.1"
/db_xref="GI:189236"

```
                     /translation="MCHQQLVISWFSLVFLASPLVAIWELKKDVYVVELDWYPDAPGE
                     MVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVLSHSLL
                     LLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWWLTTISTDLTFSVKSSR
                     GSSDPQGVTCGAATLSAERVRGDNKEYEYSVECQEDSACPAAEESLPIEVMVDAVHKL
                     KYENYTSSFFIRDIIKPDPPNNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCVQ
                     VQGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPCS"
     gene            14. .1000
                     /gene="NKSF2"
     mat_peptide     80. .997
                     /gene="NKSF2"
                     /note="G00-127-870"
                     /product="natural killer cell stimulatory factor"
BASE COUNT      692 a    478 c    535 g    613 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 9;  Length 2318;
  Best Local Similarity   87.0%;  Pred. No. 3.9e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    756 ATTCTCGGCAGGTGGAGGTCAGC 778


RESULT  39
AR082684
LOCUS       AR082684      2362 bp    DNA             PAT       01-SEP-2000
DEFINITION  Sequence 1 from patent US 5976539.
ACCESSION   AR082684
VERSION     AR082684.1  GI:10009474
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2362)
  AUTHORS   Scott,P. and Trinchieri,G.
  TITLE     Compositions and methods for use of IL-12 as an adjuvant
  JOURNAL   Patent: US 5976539-A 1 02-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .2362
                     /organism="unknown"
BASE COUNT      717 a    483 c    542 g    620 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 6;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 3.8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  40
AR122960
LOCUS       AR122960      2362 bp    DNA             PAT       16-MAY-2001
DEFINITION  Sequence 1 from patent US 6168923.
ACCESSION   AR122960
VERSION     AR122960.1  GI:14107926
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2362)
  AUTHORS   Scott,P. and Trinchieri,G.
  TITLE     Compositions and methods for use of IL-12 as an adjuvant
  JOURNAL   Patent: US 6168923-A 1 02-JAN-2001;
FEATURES             Location/Qualifiers
     source          1. .2362
                     /organism="unknown"
BASE COUNT      717 a    483 c    542 g    620 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 6;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 3.8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  41
I28325
LOCUS       I28325        2362 bp    DNA             PAT       06-FEB-1997
DEFINITION  Sequence 1 from patent US 5571515.
ACCESSION   I28325
VERSION     I28325.1  GI:1819101
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2362)
  AUTHORS   Scott,P. and Trinchieri,G.
  TITLE     Compositions and methods for use of IL-12 as an adjuvant
  JOURNAL   Patent: US 5571515-A 1 05-NOV-1996;
FEATURES             Location/Qualifiers
     source          1. .2362
                     /organism="unknown"
BASE COUNT      717 a    483 c    542 g    620 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 6;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 3.8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  42
I89770
LOCUS       I89770        2362 bp    DNA             PAT       10-AUG-1998
DEFINITION  Sequence 1 from patent US 5723127.
ACCESSION   I89770
VERSION     I89770.1  GI:3409710
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2362)
  AUTHORS   Scott,P. and Trinchieri,G.
  TITLE     Compositions and methods for use of IL-12 as an adjuvant
  JOURNAL   Patent: US 5723127-A 1 03-MAR-1998;
FEATURES             Location/Qualifiers
     source          1. .2362
                     /organism="unknown"
BASE COUNT      717 a    483 c    542 g    620 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 6;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 3.8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  43
```

```
HSAJ2967/c
LOCUS       HSAJ2967     2611 bp    mRNA            PRI       06-APR-1998
DEFINITION  Homo sapiens mRNA for cystinosin.
ACCESSION   AJ222967
VERSION     AJ222967.1  GI:3036839
KEYWORDS    cystinosin; CYSTN gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2611)
  AUTHORS   Antignac,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1997) Antignac C., U423, INSERM, Tour Lavoisier
            Hopital Necker 149 rue de Sevres Paris, 75015, FRANCE
REFERENCE   2  (bases 1 to 2611)
  AUTHORS   Town,M., Jean,G., Cherqui,S., Attard,M., Forestier,L.,
            Whitmore,S.A., Callen,D.F., Gribouval,O., Broyer,M., Bates,G.P.,
            van't Hoff,W. and Antignac,C.
  TITLE     A novel gene encoding an integral membrane protein is mutated in
            nephropathic cystinosis
  JOURNAL   Nat. Genet. 18 (4), 319-324 (1998)
  MEDLINE   98196758
FEATURES             Location/Qualifiers
     source          1. .2611
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            340. .1443
                     /gene="CTNS"
     CDS             340. .1443
                     /gene="CTNS"
                     /codon_start=1
                     /product="cystinosin"
                     /protein_id="CAA11021.1"
                     /db_xref="GI:3036840"
                     /db_xref="SPTREMBL:O60931"
                     /translation="MIRNWLTIFILFPLKLVEKCESSVSLTVPPVVKLENGSSTNVSL
                     TLRPPLNATLVITFEITFRSKNITILELPDEVVVPPGVTNSSFQVTSQNVGQLTVYLH
                     GNHSNQTGPRIRFLVIRSSAISIINQVIGWIYFVAWSISFYPQVIMNWRRKSVIGLSF
                     DFVALNLTGFVAYSVFNIGLLWVPYIKEQFLLKYPNGVNPVNSNDVFFSLHAVVLTLI
                     IIVQCCLYERGGQRVSWPAIGFLVLAWLFAFVTMIVAAVGVITWLQFLFCFSYIKLAV
                     TLVKYFPQAYMNFYYKSTEGWSIGNVLLDFTGGSFSLLQMFLQSYNNDQWTLIFGDPT
                     KFGLGVFSIVFDVVFFIQHFCLYRKRPGYDQLN"
BASE COUNT      525 a    784 c    674 g    628 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 9;  Length 2611;
  Best Local Similarity   87.0%;  Pred. No. 3.8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           || ||||| ||||||| ||||||
Db      67 GAGCCTCGCCAGGTGCCGGTCAG 45


RESULT  44
AR091393
LOCUS       AR091393     6139 bp    DNA             PAT       07-SEP-2000
DEFINITION  Sequence 7 from patent US 5994104.
ACCESSION   AR091393
VERSION     AR091393.1  GI:10018148
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6139)
  AUTHORS   Anderson,R.James, Prentice,H.Grant and MacDonald,I.Duncan.
  TITLE     Interleukin-12 fusion protein
  JOURNAL   Patent: US 5994104-A 7 30-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .6139
                     /organism="unknown"
BASE COUNT     1552 a   1538 c   1558 g   1491 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 6;  Length 6139;
  Best Local Similarity   87.0%;  Pred. No. 3.6e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db    3964 ATTCTCGGCAGGTGGAGGTCAGC 3986


RESULT  45
AE007146
LOCUS       AE007146    14698 bp    DNA             BCT       27-APR-2001
DEFINITION  Mycobacterium tuberculosis CDC1551,  section 232 of 280 of the
            complete genome.
ACCESSION   AE007146 AE000516
VERSION     AE007146.1  GI:13883176
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis CDC1551.
  ORGANISM  Mycobacterium tuberculosis CDC1551
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 14698)
  AUTHORS   Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
            Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
            Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
            Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
            Gill,J., Mikula,A. and Bishai,W.
  TITLE     Whole genome comparison of Mycobacterium tuberculosis clinical and
            laboratory strains
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 14698)
  AUTHORS   Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
            Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
            Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
            Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
            Gill,J., Mikula,A. and Bishai,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .14698
                     /organism="Mycobacterium tuberculosis CDC1551"
                     /strain="CDC1551"
                     /db_xref="taxon:83331"
                     /note="clinical strain"
     gene            complement(82. .573)
                     /gene="MT3356"
     CDS             complement(82. .573)
                     /gene="MT3356"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK47698.1"
                     /db_xref="GI:13883177"
                     /translation="MRVSGASAALVHDSLSVVNVPRRCCRPGCPHYAVATLTFVYSDS
                     TAVIGPLATAREPHSWDLCVGHAGRITAPRGWELVRHAGPLPSHPDEDDLVALADAVR
                     EGGPSAGRRHHPGGNGAPLHGFDDFPAAATGAPTGGGVLAPPEPGAGRRRGHLRVLPD
                     PAD"
     gene            588. .1115
                     /gene="MT3357"
     CDS             588. .1115
                     /gene="MT3357"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK47699.1"
```

```
                     /db_xref="GI:13883178"
                     /translation="MFNRYHRRVSDSRSSSWSRRSRGGSVARRAIRRGREMRGPLLPP
                     TVPGWRSRAERFDMAVLEAYEPIERRWQERVSQLDIAVDEIPRIAAKDPESVQWPPEV
                     IADGPIALARLIPAGVDVRGNATRARIVLFRKPIERRAKDTEELGELLHEILVAQVAI
                     YLDVDPSVIDPTIDD"
gene                 complement(1143. .1514)
                     /gene="MT3358"
CDS                  complement(1143. .1514)
                     /gene="MT3358"
                     /note="similar to GB:X68708 PID:47120; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="WhiB-related protein"
                     /protein_id="AAK47700.1"
                     /db_xref="GI:13883179"
                     /translation="MSYEHLRGVMGGTPHTTTGSATASATAVLRPHLSLVPEAPAPPE
                     EPLPPEATDQWQDRALCAQTDPEAFFPEKGGSTREAKKICMGCEVRHECLEYALAHDE
                     RFGIWGGLSERERRRLKRGII"
gene                 1814. .2809
                     /gene="MT3359"
CDS                  1814. .2809
                     /gene="MT3359"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK47701.1"
                     /db_xref="GI:13883180"
                     /translation="MKVTVLAGGVGGARFLLGVQQLLGLGQFAANSAHSDADHQLSAV
                     VNVGDDAWIHGLRVCPDLDTCMYTLGGGVDPQRGWGQRDETWHAMQELVRYGVQPDWF
                     ELGDRDLATHLVRTQMLQAGYPLSQITEALCDRWQPGARLLPATDDRCETHVVITDPV
                     DESRKAIHFQEWWVRYRAQVPTHSFAFVGAEKSSAATEAIAALADADIIMLAPSNPVV
                     SIGAILAVPGIRAALREATAPIVGYSPIIGEKPLRGMADTCLSVIGVDSTAAAVGRHY
                     GARCATGILDCWLVHDGDHAEIDGVTVRSVPLLMTDPNATAEMVRAGCDLAGVVA"
gene                 2806. .4152
                     /gene="MT3361"
CDS                  2806. .4152
                     /gene="MT3361"
                     /note="identified by match to PFAM  protein family HMM
                     PF01996"
                     /codon_start=1
                     /transl_table=11
                     /product="nitroreductase family protein"
                     /protein_id="AAK47702.1"
                     /db_xref="GI:13883181"
                     /translation="MTGPEHGSASTIEILPVIGLPEFRPGDDLSAAVAAAAPWLRDGD
                     VVVVTSKVVSKCEGRLVPAPEDPEQRDRLRRKLIEDEAVRVLARKDRTLITENRLGLV
                     QAAAGVDGSNVGRSELALLPVDPDASAATLRAGLRERLGVTVAVVITDTMGRAWRNGQ
                     TDAAVGAAGLAVLRNYAGVRDPYGNELVVTEVAVADEIAAAADLVKGKLTATPVAVVR
                     GFGVSDDGSTARQLLRPGANDLFWLGTAEALELGRQQAQLLRRSVRRFSTDPVPGDLV
                     EAAVAEALTAPAPHHTRPTRFVWLQTPAIRARLLDRMKDKWRSDLTSDGLPADAIERR
                     VARGQILYDAPEVVIPMLVPDGAHSYPDAARTDAEHTMFTVAVGAAVQALLVALAVRG
                     LGSCWIGSTIFAADLVRDELDLPVDWEPLGAIAIGYADEPSGLRDPVPAADLLILK"
gene                 4253. .4387
                     /gene="MT3362"
CDS                  4253. .4387
                     /gene="MT3362"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK47703.1"
                     /db_xref="GI:13883182"
                     /translation="MAPRLEILQEGRQDLGAVIDAAVGGALSVMLGNIPLVVPNANQL
                     "
gene                 4448. .6109
                     /gene="MT3363"
CDS                  4448. .6109
                     /gene="MT3363"
                     /note="similar to GB:M74821 SP:P25240 GB:X61122 PID:145138
                     PID:41320; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="modification methylase"
                     /protein_id="AAK47704.1"
                     /db_xref="GI:13883183"
                     /translation="MQPSHPTRPGAVIRYVGSSLDTCPMTTFAGKTAASADKVRGGYY
                     TPPAVARFLAHWVHQAGPKILEPSCGDGRILRELSAITDHAHGVELVAREAKKSRDFA
                     SVDTENLFTWLHKTQLGSWDGVAGNPPYIRFGNWASEQRDPALELMRRVGLRPTKLTN
                     AWVPFVVASTTLARDGGRVGLVVPAELLQVTYAAQLREFLLSRYREITLVTFERLVFD
                     GILQEVVLFCGVVGPGPAHIRTVRLGDANDLNALGDKDFTNESAPALLHEKEKWTKYF
                     LDPAQIRLLRGLKQSATMIRLGELADVDVGIVTGRNSFFTFTDAKAQALGLRAHCVPL
                     VSRSAQLSGLIYDEDCRACDVAGNHRTWLLDAADYPTDPALVAHITAGEAAGVHLGYK
                     CSIRKPWWSTPSLWMPDLFMLRQIHFAPRLTVNAAAATSTDTVHRVRLDPNVDPATLA
                     AVFHNSATFAFAEIMGRSYGGGILELEPREAEQLPMPPPAYGSAELAQDVDLLLKANE
                     IDKALDVVDRHVLIDGLGLSPRLVAGCRAAWLTLRDRRTKRGSRR"
gene                 complement(6169. .7248)
                     /gene="MT3364"
CDS                  complement(6169. .7248)
                     /gene="MT3364"
                     /note="similar to PID:1653786; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="mannose-1-phosphate guanyltransferase"
                     /protein_id="AAK47705.1"
                     /db_xref="GI:13883184"
                     /translation="MATHQVDAVVLVGGKGTRLRPLTLSAPKPMLPTAGLPFLTHLLS
                     RIAAAGIEHVILGTSYKPAVFEAEFGDGSALGLQIEYVTEEHPLGTGGGIANVAGKLR
                     NDTAMVFNGDVLSGADLAQLLDFHRSNRADVTLQLVRVGDPRAFGCVPTDEEDRVVAF
                     LEKTEDPPTDQINAGCYVFERNVIDRIPQGREVSVEREVFPALLADGDCKIYGYVDAS
                     YWRDMGTPEDFVRGSADLVRGIAPSPALRGHRGEQLVHDGAAVSPGALLIGGTVVGRG
                     AEIGPGTRLDGAVIFDGVRVEAGCVIERSIIGFGARIGPRALIRDGVIGDGADIGARC
                     ELLSGARVWPGVFLPDGGIRYSSDV"
gene                 complement(7250. .8155)
                     /gene="MT3365"
CDS                  complement(7250. .8155)
                     /gene="MT3365"
                     /note="similar to GB:AE000782; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glycosyl transferase"
                     /protein_id="AAK47706.1"
                     /db_xref="GI:13883185"
                     /translation="MVAVTYSPGPHLERFLASLSLATERPVSVLLADNGSTDGTPQAA
                     VQRYPNVRLLPTGANLGYGTAVNRTIAQLGEMAGDAGEPWVDDWVIVANPDVQWGPGS
                     IDALLDAASRWPRAGALGPLIRDPDGSVYPSARQMPSLIRGGMHAVLGPFWPRNPWTT
                     AYRQERLEPSERPVGWLSGSCLLVRRSAFGQVGGFDERYFMYMEDVDLGDRLGKAGWL
                     SVYVPSAEVLHHKAHSTGRDPASHLAAHHKSTYIFLADRHSGWWRAPLRWTLRGSLAL
                     RSHLMVRSSLRRSRRRKLKLVEGRH"
gene                 complement(8166. .9080)
                     /gene="MT3366"
CDS                  complement(8166. .9080)
                     /gene="MT3366"
                     /note="similar to SP:P29781 GB:X62567 PID:49013;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="dTDP-4-dehydrorhamnose reductase"
                     /protein_id="AAK47707.1"
                     /db_xref="GI:13883186"
                     /translation="MAGRSERLVITGAGGQLGSHLTAQAAREGRDMLALTSSQWDITD
                     PAAAERIIRHGDVVINCAAYTDVDGAESNEAVAYAVNATGPQHLARACARVGARLIHV
                     STDYVFDGDFGGAEPRPYEPTDETAPQGVYARSKLAGEQAVLAAFPEAAVVRTAWVYT
                     GGTGKDFVAVMRRLAAGHGRVDVVDDQTGSPTYVADLAEALLALADAGVRGRVLHAAN
                     EGVVSRFGQARAVFEECGADPQRVRPVSSAQFPRPAPRSSYSALSSRQWALAGLTPLR
                     HWRSALATALAAPANSTSIDRRLPSTRD"
gene                 9159. .10652
                     /gene="MT3367"
CDS                  9159. .10652

Query Match          72.8%;  Score 18.2;  DB 1;  Length 14698;
Best Local Similarity  87.0%;  Pred. No. 3.3e+02;
Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy       2 gatcctcgtcaggtgcaggtcag 24
           ||||| |||||||||||||| | |
Db   10155 GATCCGCGTCAGGTGCAGGACTG 10177
```

Search completed: December 26, 2001, 10:58:48
Job time: 2090 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:19:34 ; Search time 366.42 Seconds
(without alignments)
58.493 Million cell updates/sec

Title: US-09-396-196F-9
Perfect score: 25
Sequence: 1 cgatcctcgtcaggtgcaggtcagc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	25	100.0	839	22	AAI62941	Human genomic DNA
2	25	100.0	1041	20	AAX01303	E. coli biotin syn
3	25	100.0	1084	10	AAN91329	E.coli Bio B gene.
4	25	100.0	1121	7	AAN60496	Sequence encoding
5	25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
6	18.2	72.8	906	11	AAQ04701	Sequence encoding
7	18.2	72.8	979	12	AAQ12149	Cytotoxic lymphocy
8	18.2	72.8	984	18	AAT50759	Sequence of ovine
9	18.2	72.8	987	16	AAQ97484	Human interleukin-
10	18.2	72.8	987	19	AAV07273	Human Interleukin-
11	18.2	72.8	987	19	AAV07217	DNA encoding human
12	18.2	72.8	987	20	AAZ40404	Human interleukin
13	18.2	72.8	987	20	AAV83199	Interleukin (IL)-1
14	18.2	72.8	1018	15	AAQ67879	EPV 42 kD promoter
15	18.2	72.8	1018	20	AAZ08510	EPV 42 kDa/human I
16	18.2	72.8	1019	21	AAA70709	Bovine interleukin
17	18.2	72.8	1389	22	AAF76880	Human p40-IL-2 cod
18	18.2	72.8	1397	18	AAV00401	40 kDa subunit of
19	18.2	72.8	1397	18	AAT77849	40 kDa subunit of
20	18.2	72.8	1397	18	AAT97969	40 kDa subunit of
21	18.2	72.8	1397	19	AAV37767	Cytotoxic lymphocy
22	18.2	72.8	1399	21	AAZ40021	Human p40 coding s
c 23	18.2	72.8	2096	21	AAZ50892	Human receptor-ass
24	18.2	72.8	2300	21	AAA92406	Human interleukin
25	18.2	72.8	2362	13	AAQ23586	Natural killer cel
26	18.2	72.8	2362	17	AAT48098	Human interleukin-
27	18.2	72.8	2362	19	AAV42535	Human interleukin-
28	18.2	72.8	2362	19	AAV15227	Heterodimeric huma
29	18.2	72.8	2362	21	AAZ28862	Human interleukin-
30	18.2	72.8	2362	22	AAF56280	IL-12 vaccine adju
31	18.2	72.8	5686	20	AAZ40415	Plasmid pIN1143 co
32	18.2	72.8	6139	21	AAZ40023	Interleukin-12 fus
33	17.8	71.2	316	22	AAF98655	Human ovarian canc
c 34	17.2	68.8	586	21	AAC75563	Human ORFX ORF1118
c 35	17.2	68.8	6372	22	AAH43851	Human CLASP-7 enco
c 36	16.8	67.2	5269	21	AAC76615	Human ORFX ORF2170
37	16.6	66.4	34	16	AAQ75832	Sense primer to am
38	16.6	66.4	220	22	AAF28005	Human TR13 coding
39	16.6	66.4	220	22	AAF28044	Human TR13 coding
40	16.6	66.4	625	21	AAC98199	Human colon cancer
41	16.6	66.4	651	21	AAF14300	Aspergillus oryzae
42	16.6	66.4	681	22	AAF71312	Corynebacterium gl
43	16.6	66.4	924	20	AAX18177	Canine mature CLAF
44	16.6	66.4	990	19	AAV13816	Canine IL-12 P40 s
45	16.6	66.4	990	19	AAV13817	Canine IL-12 P40 s
46	16.6	66.4	990	20	AAX35615	Nucleic acid encod
47	16.6	66.4	990	20	AAX35625	Nucleic acid encod
48	16.6	66.4	990	20	AAX03162	Nucleic acid encod
49	16.6	66.4	990	20	AAX03152	Nucleic acid encod
50	16.6	66.4	1143	22	AAH66315	C glutamicum codin
c 51	16.6	66.4	1485	21	AAA37076	Human PRO1306 (UNQ
c 52	16.6	66.4	1485	22	AAF54342	DNA encoding prote
53	16.6	66.4	1717	20	AAZ41991	Human endometrium
c 54	16.6	66.4	1731	22	AAI58050	Human polynucleoti
c 55	16.6	66.4	1743	22	AAI59836	Human polynucleoti
56	16.6	66.4	2154	20	AAX18176	Canine full length
57	16.6	66.4	2554	22	AAF27997	Human TR13 recepto
58	16.6	66.4	2610	22	AAF89775	Nucleotide sequenc
59	16.6	66.4	2733	22	AAF89774	Nucleotide sequenc
60	16.6	66.4	3280	21	AAA95442	Human CASB619 prot
61	16.6	66.4	3331	22	AAF89765	Nucleotide sequenc
62	16.6	66.4	3334	22	AAF28030	Human TR13 recepto
c 63	16.6	66.4	3375	22	AAH17675	Human cDNA sequenc
c 64	16.6	66.4	3449	22	AAC66521	Human immune syste
65	16.6	66.4	3501	22	AAS21262	Human cDNA sequenc
c 66	16.6	66.4	11272	20	AAX89842	Rat cubilin protei
67	16.6	66.4	349980	22	AAH68528	C glutamicum codin
c 68	16.2	64.8	403	21	AAF12855	Aspergillus oryzae
69	16.2	64.8	669	21	AAF14275	Aspergillus oryzae
c 70	16.2	64.8	3509	13	AAQ20664	Sucrose phosphate
c 71	16.2	64.8	3509	13	AAQ29926	SPS gene. Zea may
c 72	16.2	64.8	3509	18	AAT95847	Maize sucrose phos
c 73	16.2	64.8	3509	18	AAT66229	Sucrose phosphate
c 74	16.2	64.8	6901	15	AAQ73703	Rice sucrose phosp
c 75	16	64.0	181	22	AAI21230	Probe #11163 for g
c 76	16	64.0	181	22	AAI46502	Probe #15188 used
c 77	16	64.0	181	22	AAI06939	Probe #6930 used t
c 78	16	64.0	450	22	AAI12043	Probe #1976 for ge
c 79	16	64.0	450	22	AAI33378	Probe #2064 used t
c 80	16	64.0	450	22	AAI01968	Probe #1959 used t
c 81	16	64.0	465	22	AAI12026	Probe #1959 for ge
c 82	16	64.0	465	22	AAI33361	Probe #2047 used t
c 83	16	64.0	465	22	AAI01953	Probe #1944 used t
84	16	64.0	489	22	AAI14805	Probe #4738 for ge

	85	16	64.0	489	22	AAI36163	Probe #4849 used t
	86	16	64.0	489	22	AAI04596	Probe #4587 used t
	87	16	64.0	603	22	AAF71867	Corynebacterium gl
	88	16	64.0	603	22	AAF72111	Corynebacterium gl
	89	16	64.0	603	22	AAF72132	Corynebacterium gl
c	90	16	64.0	717	21	AAA79473	Eucalyptus grandis
c	91	16	64.0	2073	21	AAX89443	Entire soybean dih
	92	16	64.0	2235	22	AAH66226	C glutamicum codin
	93	16	64.0	2358	22	AAF71865	Corynebacterium gl
	94	16	64.0	2358	22	AAF72109	Corynebacterium gl
	95	16	64.0	2358	22	AAF72130	Corynebacterium gl
c	96	16	64.0	349980	22	AAH68527	C glutamicum codin
	97	15.8	63.2	442	21	AAC16888	Human secreted pro
	98	15.8	63.2	1155	22	AAH66024	C glutamicum codin
c	99	15.8	63.2	1183	20	AAZ31920	Human helicase, Re
c	100	15.8	63.2	1749	22	AAF87638	Human RecQ5 gamma

ALIGNMENTS

```
RESULT   1
AAI62941
ID   AAI62941 standard; DNA; 839 BP.
XX
AC   AAI62941;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human genomic DNA SEQ ID NO 269.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant;  immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200155449-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01346.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUL-2000; 2000US-0216880.
PR   14-JUL-2000; 2000US-0218290.
PR   14-AUG-2000; 2000US-0225447.
PR   01-SEP-2000; 2000US-0229343.
PR   06-SEP-2000; 2000US-0230437.
PR   08-SEP-2000; 2000US-0231243.
PR   25-SEP-2000; 2000US-0234997.
PR   29-SEP-2000; 2000US-0236367.
PR   13-OCT-2000; 2000US-0239937.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246528.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249265.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-476225/51.
XX
PT   Novel plasma membrane associated proteins useful for diagnosing,
PT   treating, preventing and/or prognosing disorders related to the
PT   proteins, including cancer, immune response and neuronal disorders
XX
PS   Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (AAI62752-AAI62961) and proteins
CC   (AAM42347-AAM42415) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;


  Query Match           100.0%;  Score 25;  DB 22;  Length 839;
  Best Local Similarity  100.0%;  Pred. No. 0.033;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db     137 cgatcctcgtcaggtgcaggtcagc 161


RESULT   2
AAX01303
ID   AAX01303 standard; DNA; 1041 BP.
XX
AC   AAX01303;
XX
DT   12-APR-1999  (first entry)
XX
DE   E. coli biotin synthetase (BioB) coding sequence.
XX
KW   DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW   biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS   Escherichia coli.
XX
PN   US5869719-A.
XX
PD   09-FEB-1999.
XX
PF   30-APR-1997;   97US-0846338.
XX
PR   30-APR-1997;   97US-0846338.
PR   08-MAR-1995;   95US-0401068.
XX
```

```
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Patton DA;
XX
DR   WPI; 1999-152902/13.
DR   P-PSDB; AAW73906.
XX
PT   Transgenic plants with high biotin levels - transformed with DNA
PT   encoding di:amino-pelargonic acid amino-transferase or biotin
PT   synthase
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.033;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    105 cgatcctcgtcaggtgcaggtcagc 129


RESULT   3
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA;
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 0.034;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    128 cgatcctcgtcaggtgcaggtcagc 152


RESULT   4
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzyme -
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 0.034;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcagc 25
          |||||||||||||||||||||||||
Db    146 cgatcctcgtcaggtgcaggtcagc 170


RESULT   5
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
```

```
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry) .
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag= a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag= b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag= c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag= d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag= e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag= f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag= g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag= h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag= i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag= j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag= k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag= 1
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag= m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag= n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                   7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag= o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag= p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag= q
FT                   /standard_name= "rho-independent transcriptional
FT                                   terminator"
FT   stem_loop       5583..5605
FT                   /*tag= r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;   93WO-EP02688.
XX
PR   02-OCT-1992;   92CH-0003124.
PR   15-JUL-1993;   93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.039;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db     221 cgatcctcgtcaggtgcaggtcagc 245


RESULT   6
AAQ04701
ID   AAQ04701 standard; DNA; 906 BP.
XX
AC   AAQ04701;
```

```
XX
DT   11-OCT-1990  (first entry)
XX
DE   Sequence encoding natural killer cell stimulatory factor (NKSF).
XX
KW   Natural killer cell stimulatory factor; NKSF; cancer;
KW   leukopaenia; anaemia; radiation; bone marrow transplantation; ds.
XX
OS   Homo sapiens.
XX
PN   WO9005147-A.
XX
PD   17-MAY-1990.
XX
PF   09-NOV-1989;   89WO-0005027.
XX
PR   10-NOV-1988;   88US-0269945.
XX
PA   (GENE-) GENETICS INST INC.
PA   (WIST-) WISTAR INST.
XX
PI   Trinchieri G,  Perussia B,  Kobayashi M,  Clark SC,  Wong GG;
XX
DR   WPI; 1990-178800/23.
DR   P-PSDB; AAR05171.
XX
PT   New natural killer cell stimulatory factor -
PT   useful for treating cancer and eg. leukopenia, bacterial and
PT   viral infections, anaemia and B- or T-cell deficiencies.
XX
PS   Claim 16; Page 60; 69pp; English.
XX
CC   Natural killer cell stimulatory factor (NKSF) is useful in
CC   treating cancer and other disease states which respond to NK cell
CC   activity. Such states may arise from immune or haematopoietic-cell
CC   deficiency following a bone marrow translpantation, exposure to
CC   radiation and disease.
XX
SQ   Sequence 906 BP; 244 A; 220 C; 238 G; 204 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 11;  Length 906;
  Best Local Similarity   87.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     775 attctcggcaggtggaggtcagc 797


RESULT   7
AAQ12149
ID   AAQ12149 standard; DNA; 979 BP.
XX
AC   AAQ12149;
XX
DT   10-SEP-1991  (first entry)
XX
DE   Cytotoxic lymphocyte maturation factor (40 kD subunit) gene.
XX
KW   Cytotoxic lymphocyte maturation factor; CLMF; probe; antitumour;
KW   monoclonal antibodies; T cell; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..979
FT                   /*tag=  a
FT                   /product= CLMF
FT   sig_peptide     41..106
FT                   /*tag=  b
FT                   /note= "hydrophobic signal peptide"
FT   mat_peptide     107..979
FT                   /*tag=  c
XX
PN   EP433827-A.
XX
PD   26-JUN-1991.
XX
PF   09-DEC-1990;   90EP-0123670.
XX
PR   27-AUG-1990;   90US-0572284.
PR   22-DEC-1989;   89US-0455708.
PR   09-MAY-1990;   90US-0520935.
XX
PA   (HOFF ) HOFFMANN-LA ROCHE AG.
XX
PI   Chizzonite RA,  Gately MK,  Gubler UA,  Hulmes JD,  Pan YCE;
PI   Podlaski FJ,  Stern AS;
XX
DR   WPI; 1991-186715/26.
DR   P-PSDB; AAR12493.
XX
PT   New cytotoxic lymphocyte maturation factor - used as antitumour
PT   agent and monoclonal antibodies used in transplantation to block
PT   proliferation and activation of cytotoxic T cells
XX
PS   Disclosure; Fig 25 (a-c); 90pp; English.
XX
CC   The natural CLMF protein is a 75 kD heterodimer comprising two
CC   polypeptide subunits, a 40 kD and a 35 kD subunit (AAQ12150).
CC   The subunits are bonded together via one or more disulfide bonds.
CC   The CLMF can interact with IL-2 to synergistically induce the
CC   cytolytic activity of Lymphokine Activated Killer (LAK) cells.  It is
CC   also capable of stimulating T-cell growth.  It can be used for the
CC   prepn. of LAK- and T-cell activator and antitumour compsns. and for
CC   stimulating Natural Killer (NK) cells.
XX
SQ   Sequence 979 BP; 261 A; 245 C; 257 G; 216 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 12;  Length 979;
  Best Local Similarity   87.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     783 attctcggcaggtggaggtcagc 805


RESULT   8
AAT50759
ID   AAT50759 standard; cDNA; 984 BP.
XX
AC   AAT50759;
XX
DT   24-SEP-1997  (first entry)
XX
DE   Sequence of ovine IL-12 40 kD sununit cDNA.
XX
KW   Cytokine; ovine; sheep; interleukin-5; interleukin-12; IL-5; IL-12;
KW   livestock; cow; stress; transport; vaccine adjuvant; veterinary;
KW   cancer; immunosuppression; allergy; reproductive system; growth;
KW   early maturity; antibody; diagnosis; immunopotentiator;
KW   early haematopoietic progenitor cell; cytotoxic cell; thymocyte;
KW   secretion; IgM; IgA; bacterial endotoxin; gamma-interferon; ss.
XX
OS   Ovis ammon aries.
XX
PN   WO9700321-A1.
XX
PD   03-JAN-1997.
XX
PF   14-JUN-1996;   96WO-AU00360.
XX
PR   27-OCT-1995;   95AU-0006244.
```

```
PR   14-JUN-1995;   95AU-0003502.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI   Seow H,  Wood P;
XX
DR   WPI; 1997-077528/07.
DR   P-PSDB; AAW08483.
XX
PT   Nucleic acid encoding ovine interleukin-5 or -12 - used as vaccine
PT   adjuvants and to treat or prevent microbial infections in livestock
XX
PS   Claim 10; Page 49-51; 78pp; English.
XX
CC   The sequences given in AAT50757-59 encode ovine interleukin-12 (IL-12)
CC   35 kD subunit (partial and full length sequence) and the 40 kD subunit.
CC   Ovine IL-5 or IL-12 are used to treat and/or prevent infections in
CC   livestock (esp. cows and sheep), particularly where the animals are
CC   stressed, e.g. during transport. IL-5 and IL-12 can also be used as
CC   adjuvants in vaccines for veterinary use (partic. weakly immunogenic
CC   subunit or synthetic peptide vaccines). They may also be used to
CC   treat cancer, immunosuppression and allergy, to enhance/suppress the
CC   reproductive system and to promote growth or early maturity.  Optionally
CC   interleukin can be delivered from constructs or delivery cells and
CC   antibodies are useful in enzyme immunoassays for rapid diagnosis of
CC   infection. The interleukins are immunopotentiators, especially IL-5
CC   promotes growth of early haematopoietic progenitor cells and generation
CC   of cytotoxic cells from thymocytes, also it stimulates production and
CC   secretion of IgM and IgA (in synergism with bacterial endotoxin).
CC   IL-12 induces production of gamma-interferon by, and proliferation
CC   of, T and NK cells and increases the (non-)specific cytolytic
CC   lymphocyte response.  The genetic constructs can also be used for
CC   in vitro production of IL-5 or -12.
XX
SQ   Sequence 984 BP; 273 A; 243 C; 253 G; 215 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 18;  Length 984;
  Best Local Similarity   87.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     746 attctcggcaggtggaggtcagc 768


RESULT   9
AAQ97484
ID   AAQ97484 standard; cDNA; 987 BP.
XX
AC   AAQ97484;
XX
DT   22-DEC-1995  (first entry)
XX
DE   Human interleukin-12 40 kD subunit.
XX
KW   Interleukin-12; IL-12; cytokine; gamma-interferon; T cell;
KW   natural killer cell; graft vs. host disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..987
FT                   /*tag=  a
XX
PN   WO9519786-A.
XX
PD   27-JUL-1995.
XX
PF   20-JAN-1995;   95WO-US00879.
XX
PR   21-JAN-1994;   94US-0186529.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (GEMY ) GENETICS INST INC.
XX
PI   Sykes M,  Wolf SF;
XX
DR   WPI; 1995-269273/35.
DR   P-PSDB; AAR79186.
XX
PT   Control of graft versus host disease using interleukin 12 - partic.
PT   admin. at the time of bone marrow transplant
XX
PS   Disclosure; Page 21-22; 32pp; English.
XX
CC   AAQ97484 and AAQ97485 are from human lymphoblast cell line RMPI 8866.
CC   IL- 12 is a heterodimeric cytokine which was originally identified as a
CC   factor which induces gamma-interferon from T cells and natural
CC   killer cells as set forth in PCT/US91/06332, published April 2 1992,
CC   and incorporated by reference. EP 433827, published June 26 1991,
CC   disclosed IL-12 as a cytotoxin lymphocyte maturation factor. In the
CC   patent, IL-12 is defined as a heterodimeric glycoprotein comprised
CC   of two covalently linked subunits having mol. wts of about 40 kD and
CC   35 kD. The two subunits may be disulfide-bonded. IL-12 is
CC   administered to prevent or ameliorate GVHD. Dosage is 1 ng-0.1 mg
CC   (esp. 1.0 ng-1 microg)/kg daily for 3 days, the first dose given
CC   1 hr. after transplant. IL-12 can be used as the native heterodimer
CC   or as subunits.
XX
SQ   Sequence 987 BP; 264 A; 243 C; 259 G; 221 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 16;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 attctcggcaggtggaggtcagc 765


RESULT  10
AAV07273
ID   AAV07273 standard; cDNA; 987 BP.
XX
AC   AAV07273;
XX
DT   25-SEP-1998  (first entry)
XX
DE   Human Interleukin-12 p40 gene.
XX
KW   IL-12 subunit; expression construct; treatment; asthma; microbial
KW   infection; viral infection; cancer; Human; Interleukin; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..987
FT                   /*tag=  a
FT                   /product= "IL-12 p40"
XX
PN   WO9817689-A2.
XX
PD   30-APR-1998.
XX
PF   10-OCT-1997;   97WO-US18779.
XX
PR   18-OCT-1996;   96US-0028676.
XX
PA   (GENE-) GENEMEDICINE INC.
XX
PI   Deshpande D,  Freimark B,  Nordstrom J;
XX
DR   WPI; 1998-261428/23.
DR   P-PSDB; AAW51311.
```

```
XX
PT   Constructs for expression of interleukin-12 sub-units  - are used
PT   for delivery of IL-12 sub-units for treating e.g. asthma, microbial
PT   or viral infections and certain cancers
XX
PS   Claim 1; Page 34; 80pp; English.
XX
CC   The IL-12 subunits AAV07273 and AAV07274 can be expressed using an
CC   appropriate expression construct engineered into a plasmid.  The
CC   plasmids can provide for efficient expression of IL-12 subunits.  The
CC   products can be used for the treatment of asthma, microbial and viral
CC   infections and certain cancers.
XX
SQ   Sequence 987 BP; 264 A; 243 C; 259 G; 221 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 19;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    743 attctcggcaggtggaggtcagc 765


RESULT  11
AAV07217
ID   AAV07217 standard; DNA; 987 BP.
XX
AC   AAV07217;
XX
DT   12-AUG-1998  (first entry)
XX
DE   DNA encoding human interleukin-12 40kDa subunit.
XX
KW   Interleukin-12; IL-12; graft-versus-host disease; GVHD.
KW   bone marrow transplantation; graft-versus-leukemia; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..987
FT                   /*tag=  a
FT                   /product= Interleukin-12_40kDa_subunit
XX
PN   US5756085-A.
XX
PD   26-MAY-1998.
XX
PF   30-APR-1996;   96US-0640386.
XX
PR   30-APR-1996;   96US-0640386.
PR   21-JAN-1994;   94US-0186529.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (GEMY ) GENETICS INST INC.
XX
PI   Sykes M,  Wolf SF;
XX
DR   WPI; 1998-321459/28.
DR   P-PSDB; AAW51221.
XX
PT   Therapy of graft-versus-host disease  - by administration of
PT   interleukin-12 to maintain and/or promote graft-versus-leukaemia
PT   effects
XX
PS   Disclosure; Columns 17-20; 20pp; English.
XX
CC   The invention relates to a method for the treatment or prevention
CC   of graft-versus-host disease (GVHD) which comprises the administration
CC   of interleukin-12 (IL-12) at the time of bone-marrow transplantation to
CC   maintain or promote graft-versus-leukaemia effects. Preferably, dosage is
CC   0.001-100 (preferably 1-100) mu g/kg for 3 days starting on the day of
CC   the bone marrow transplant. IL-12 is administered daily until there is
CC   an improvement in the condition. Administration is oral, cutaneous,
CC   subcutaneous, intravenous or by inhalation preferably subcutaneous, or
CC   intravenous. The present sequence encodes the human interleukin-12
CC   40kDa subunit.
XX
SQ   Sequence 987 BP; 264 A; 243 C; 259 G; 221 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 19;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    743 attctcggcaggtggaggtcagc 765


RESULT  12
AAZ40404
ID   AAZ40404 standard; DNA; 987 BP.
XX
AC   AAZ40404;
XX
DT   15-FEB-2000  (first entry)
XX
DE   Human interleukin 12 p40 subunit wild type coding sequence.
XX
KW   Wild type; human; interferon-alpha; plasmid; cytomegalovirus; CMV;
KW   promoter; growth hormone; untranslated region; UTR; mammal; disease;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9947678-A2.
XX
PD   23-SEP-1999.
XX
PF   12-MAR-1999;   99WO-US05394.
XX
PR   19-MAR-1998;   98US-0078654.
XX
PA   (GENE-) GENEMEDICINE INC.
XX
PI   Nordstrom J,  Pericle F,  Rolland A,  Ralston R;
XX
DR   WPI; 1999-562116/47.
DR   P-PSDB; AAY42487.
XX
PT   New plasmids containing an interferon-alpha coding sequence, used for
PT   the treatment of a mammalian condition or disease, particularly cancer
PT   -
XX
PS   Claim 11; Page 43-45; 137pp; English.
XX
CC   This sequence represents the wild type coding region for the human
CC   interleukin 12 p40 subunit. The invention relates to a novel plasmid
CC   comprising a cytomegalovirus (CMV) promoter transcriptionally linked
CC   with an interferon alpha (IFN-alpha) coding sequence, and a growth
CC   hormone 3'-untranslated region (UTR).  The plasmids can be used for
CC   treating a mammalian condition or disease, particularly cancer.
XX
SQ   Sequence 987 BP; 264 A; 243 C; 259 G; 221 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 20;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    743 attctcggcaggtggaggtcagc 765
```

```
RESULT  13
AAV83199
ID   AAV83199 standard; DNA; 987 BP.
XX
AC   AAV83199;
XX
DT   02-MAR-1999  (first entry)
XX
DE   Interleukin (IL)-12B cDNA.
XX
KW   Mouse; immunodeficient; pathogen; tumour; lymphocyte; antigen;
KW   immunomodulator; vector; vaccine; cancer; HIV; Leishmania;
KW   Mycobacterium; Listeria; Plasmodium; retrovirus; evaluation;
KW   human immunodeficiency virus; ds.
XX
OS   Homo sapiens.
XX
PN   WO9844788-A2.
XX
PD   15-OCT-1998.
XX
PF   09-APR-1998;   98WO-US06944.
XX
PR   09-DEC-1997;   97US-0069163.
PR   09-APR-1997;   97US-0838702.
PR   01-MAY-1997;   97US-0848760.
XX
PA   (CHAN/) CHANG L.
XX
PI   Chang L;
XX
DR   WPI; 1999-024005/02.
XX
PT   Use of immunodeficient mice comprising human cells - particularly
PT   SCTD/beige mice comprising human immune cells for evaluating
PT   vaccines against cancers or human pathogens, e.g. HIV
XX
PS   Example 10(c)(i); Page 110; 154pp; English.
XX
CC   Immunodeficient mice comprising human cells can be used for
CC   exposure to human pathogens and/or their components or human
CC   tumour cells and human peripheral blood lymphocytes. Also claimed
CC   is a vaccine comprising a cell modified to express an antigen and
CC   an immune-modulating protein, this is preferably an expression
CC   vector comprising a polynucleotide sequence that encodes the
CC   antigen and immune-modulating protein.  Such vectors can be used
CC   to treat a subject having a tumour by transferring the expression
CC   vector into the tumour so that the antigen and the immune-modulator
CC   are expressed by at least the tumour. The methods can be used for
CC   producing and evaluating vaccines including cancer vaccines and
CC   vaccines directed against human pathogens, e.g. HIV, Leishmania,
CC   Mycobacterium, Listeria or Plasmodium.  Two primers (AAV83197,
CC   AAV83198) were used to amplify the IL-12B for insertion into
CC   expression vectors of the invention.
XX
SQ   Sequence 987 BP; 264 A; 245 C; 257 G; 221 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 20;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 attctcggcaggtggaggtcagc 765


RESULT  14
AAQ67879
ID   AAQ67879 standard; DNA; 1018 BP.
XX
AC   AAQ67879;
XX
DT   23-MAR-1995  (first entry)
XX
DE   EPV 42 kD promoter/Human IL-12 p40 subunit expression cassette.
XX
KW   Polymerase chain reaction; primer; amplify; NYVAC; ALVAC; recombinant;
KW   murine; interleukin-2; IL-2; pRW825; pmut-1; pBS-SK; pMM151; TK vector;
KW   plasmid; vaccinia; H6 promoter; amplify; primer; antigenic response;
KW   polymerase chain reaction; poxvirus; pSD542; immunological response;
KW   pathogen; human; interferon; IFN; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..31
FT                   /*tag=  a
FT                   /note= "EPV 42 kD promoter"
FT   CDS             32..1018
FT                   /*tag=  b
FT                   /product= Human IL-12 p40
XX
PN   WO9416716-A.
XX
PD   04-AUG-1994.
XX
PF   21-JAN-1994;   94WO-US00888.
XX
PR   21-JAN-1993;   93US-0007115.
PR   19-JAN-1994;   94US-0184009.
XX
PA   (VIRO-) VIROGENETICS CORP.
XX
PI   Cox WI,  Paoletti E,  Tartaglia J;
XX
DR   WPI; 1994-263767/32.
XX
PT   Attenuated recombinant virus used for cancer therapy - comprises
PT   DNA encoding cytokine and/or tumour associated antigen
XX
PS   Example 27; Fig 33; 232pp; English.
XX
CC   The sequences given in AAQ67879-80 are expression cassettes which
CC   contain the p40 and p35 subunits of human interleukin-12 (IL-12) under
CC   the control of the entomopox virus (EPV) promoter and the vaccinia E3L
CC   promoter, respectively.  These sequences were used in the construction
CC   of an ALVAC-based recombinant virus containing the genes for the two
CC   subunits of human IL-12.  The p40 subunit was amplified using primers
CC   JP202 and JP189 and placed under the control of the entomopox virus
CC   (EPV) 42 kD promoter in plasmid PBSHIL12p40II.  The p35 subunit was
CC   amplified using primers JP186 and JP201 and placed under the control
CC   of the vaccinia E3L promoter using primers JP218 and JP220.  The
CC   expression cassette was isolated and ligated with pBSSK+ generating
CC   plasmid PBSHIL12p35II.  The p35 and p40 expression cassettes were
CC   excised and assembled in pBSSK+ to give PBSHIL12 in which the two
CC   expression cassettes are orientated head-to-head relative to each other.
CC   The combination casette containing the poxvirus promoted genes for both
CC   IL-12 subunits was excised from PBSHIL12 and cloned in to ALVAC C6
CC   vector plasmid pC6L to give pC6HIL12.  Recombination between this donor
CC   plasmid and ALVAC rescuing virus produced recombinant virus ALVAC+IL12,
CC   which contains both the human IL-12 genes in the C6 locus.  The resulting
CC   virus may be used in a composition for inducing an antigenic or
CC   immunological response, ie. for immunisation against pathogens.
XX
SQ   Sequence 1018 BP; 283 A; 245 C; 260 G; 230 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 15;  Length 1018;
  Best Local Similarity   87.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     774 attctcggcaggtggaggtcagc 796
```

```
RESULT  15
AAZ08510
ID   AAZ08510 standard; DNA; 1018 BP.
XX
AC   AAZ08510;
XX
DT   19-OCT-1999  (first entry)
XX
DE   EPV 42 kDa/human IL-12 P40 expression cassette.
XX
KW   Attenuated recombinant virus; cytokine; tumour associated antigen;
KW   NYVAC recombinant virus; ALVAC recombinant virus; gene therapy; rabies;
KW   cancer; tumour necrosis factor; nuclear phosphoprotein; p53; IL-2; GMCSF;
KW   interleukin; interferon; IFN-gamma; IL-4; melanoma associated antigen;
KW   carcinoembryonic antigen; immunisation; antigenic; poxvirus; influenza;
KW   immunological response; immunotherapy; vaccine; Newcastle Disease; ss.
XX
OS   Synthetic.
OS   Entomopox virus.
OS   Homo sapiens.
XX
PN   US5942235-A.
XX
PD   24-AUG-1999.
XX
PF   02-JUN-1995;   95US-0458356.
XX
PR   02-JUN-1995;   95US-0458356.
PR   24-DEC-1981;   81US-0334456.
PR   08-DEC-1982;   82US-0446824.
PR   19-JUN-1984;   84US-0622135.
PR   27-AUG-1987;   87US-0090209.
PR   28-AUG-1987;   87US-0090711.
PR   20-OCT-1987;   87US-0110335.
PR   25-APR-1988;   88US-0186054.
PR   23-AUG-1988;   88US-0234390.
PR   08-MAR-1989;   89US-0320471.
PR   14-FEB-1990;   90US-0478179.
PR   14-JUN-1990;   90US-0537882.
PR   14-JUN-1990;   90US-0537890.
PR   07-JAN-1991;   91US-0638080.
PR   07-MAR-1991;   91US-0666056.
PR   11-JUN-1991;   91US-0713967.
PR   16-DEC-1991;   91US-0805567.
PR   03-MAR-1992;   92US-0847977.
PR   06-MAR-1992;   92US-0847951.
PR   04-MAY-1992;   92US-0881995.
PR   22-JUL-1992;   92US-0918278.
PR   20-JAN-1993;   93US-0007115.
PR   19-JAN-1994;   94US-0184009.
PR   14-APR-1994;   94US-0228926.
PR   13-SEP-1994;   94US-0306259.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Paoletti E;
XX
DR   WPI; 1999-493494/41.
XX
PT   Recombinant poxviruses comprising exogenous DNA encoding antigenic
PT   determinants useful in immunotherapy to immunize against cancers and
PT   other diseases such as influenza, Newcastle Disease and rabies
XX
PS   Example 27; Fig 33; 163pp; English.
XX
CC   The present invention describes a recombinant poxvirus (I), comprising
CC   exogenous DNA encoding an antigenic determinant of a pathogen which is
CC   then expressed in vivo in infected host cells after administration to a
CC   patient and therefore induces an immunological response. (I) may be used
CC   to vaccinate patients against a wide range of diseases and disorders
CC   depending on the type of antigen encoded by the exogenous DNA. (I) may
CC   be used to vaccinate against diseases such as rabies, influenza and
CC   Newcastle Disease. It is particularly useful for immunising against
CC   cancers. The poxvirus (I) also provides a means of manipulating
CC   lymphocytes and tumour cells for use in cell-based immunotherapeutic
CC   modalities for cancer. (I) also have enhanced safety compared to
CC   unattenuated viruses (attenuation reduces the virulence of the viruses)
CC   and known recombinant poxvirus vaccines. This increased level of safety
CC   reduces the possibility of a 'runaway' infection in the host and reduces
CC   the chance of transmission from vaccinated to unvaccinated individuals
CC   and contamination of the environment. The present sequence represents a
CC   EPV 42 kDa/human IL-12 P40 expression cassette used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 1018 BP; 283 A; 245 C; 260 G; 230 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 20;  Length 1018;
  Best Local Similarity   87.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     774 attctcggcaggtggaggtcagc 796


RESULT  16
AAA70709
ID   AAA70709 standard; DNA; 1019 BP.
XX
AC   AAA70709;
XX
DT   15-DEC-2000  (first entry)
XX
DE   Bovine interleukin-12 p40 subunit gene.
XX
KW   Antiinfective; bovine; interleukin 12; IL12; heterodimer; infection; ss.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   CDS             12..995
FT                   /*tag=  a
FT                   /product= "IL-12 p40 subunit"
XX
PN   JP2000157274-A.
XX
PD   13-JUN-2000.
XX
PF   20-NOV-1998;   98JP-0331052.
XX
PR   20-NOV-1998;   98JP-0331052.
XX
PA   (DAUC ) DAIICHI PHARM CO LTD.
XX
DR   WPI; 2000-468201/41.
DR   P-PSDB; AAB15391.
XX
PT   Preparation of bovine interleukin 12 comprising using a vector
PT   containing the gene encoding it, useful for the prevention and
PT   treatment of chronic and opportunistic infections -
XX
PS   Example 1; Page 8-9; 11pp; Japanese.
XX
CC   The invention relates to a method for the preparation of bovine
CC   interleukin 12 (BoIL12). BoIL12 consists of a heterodimer of p35
CC   and p40 subunits. BoIL12 is produced by introducing into a host cell,
CC   recombinant vectors containing genes encoding the p35 and p40 subunits.
CC   This sequence represents the coding region for the bovine IL12 p40
CC   subunit. Interleukin 12 is used for the prevention and treatment of
CC   chronic infections and opportunistic infections.
XX
SQ   Sequence 1019 BP; 273 A; 248 C; 271 G; 227 T; 0 other;
```

```
  Query Match             72.8%;  Score 18.2;  DB 21;  Length 1019;
  Best Local Similarity   87.0%;  Pred. No. 38;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     757 attctcggcaggtggaggtcagc 779


RESULT  17
AAF76880
ID   AAF76880 standard; DNA; 1389 BP.
XX
AC   AAF76880;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Human p40-IL-2 coding sequence.
XX
KW   Human; p40-IL-2; interleukin-2; IL-2; antiviral; antibacterial;
KW   antifungal; cytostatic; antiparasitic; vaccine; gene therapy;
KW   fusion protein; protein complex; cytokine; immunoglobulin; cancer;
KW   infection; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200110912-A1.
XX
PD   15-FEB-2001.
XX
PF   09-AUG-2000; 2000WO-US21715.
XX
PR   09-AUG-1999;   99US-0147924.
XX
PA   (LEXI-) LEXIGEN PHARM CORP.
XX
PI   Gillies SD,  Lo KM;
XX
DR   WPI; 2001-191523/19.
XX
PT   Novel multifunctional fusion protein or protein complexes useful for
PT   treating cancer and viral infections, comprise two different cytokine
PT   molecules and a targeting group -
XX
PS   Example 1; Page 75-76; 90pp; English.
XX
CC   The present sequence is given in a specification relating to
CC   multifunctional fusion proteins and protein complexes comprising at
CC   least two different cytokine molecules and a targeting group such as an
CC   immunoglobulin. The fusion proteins and protein complexes are useful for
CC   treating cancers, and bacterial, viral, fungal and parasitic infections.
CC   They are also useful in gene therapy and as vaccines to prevent or treat
CC   various diseases.
XX
SQ   Sequence 1389 BP; 413 A; 325 C; 327 G; 324 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 22;  Length 1389;
  Best Local Similarity   87.0%;  Pred. No. 39;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 attctcggcaggtggaggtcagc 765


RESULT  18
AAV00401
ID   AAV00401 standard; DNA; 1397 BP.
XX
AC   AAV00401;
XX
DT   18-MAR-1998  (first entry)
XX
DE   40 kDa subunit of human cytotoxic lymphocyte maturation factor DNA.
XX
KW   Cytotoxic lymphocyte maturation factor; CLMF; 40 kDa subunit;
KW   interleukin-2; T-cell growth; antibody; lymphokine activated killer cell;
KW   CLMF bioactivity inhibition; affinity ligand; transplantation;
KW   cytotoxic T cell proliferation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1027
FT                   /*tag=  a
XX
PN   EP790309-A1.
XX
PD   20-AUG-1997.
XX
PF   09-DEC-1990;   90EP-0123670.
XX
PR   27-AUG-1990;   90US-0572284.
PR   22-DEC-1989;   89US-0455708.
PR   09-MAY-1990;   90US-0520935.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Chizzonite RA,  Gately MK,  Gubler UA,  Hulmes J,  Pan YE;
PI   Podlanski FJ, Stern AS;
XX
DR   WPI; 1997-404734/38.
DR   P-PSDB; AAW36941.
XX
PT   Sub-unit of cytotoxic lymphocyte maturation factor protein - useful
PT   for producing CLMF protein and as an antitumour agent
XX
PS   Disclosure; Fig 25; 83pp; English.
XX
CC   The present DNA sequence encodes a novel 40 kDa subunit of cytotoxic
CC   lymphocyte maturation factor (CLMF) protein. The natural CLMF protein is
CC   a 75 kDa heterodimer, where the 2 subunits (see AAV00402 for the 35 kDa
CC   subunit) are bonded together via one or more disulphide bonds. CLMF
CC   protein is produced by a human B lyphoblastoid cell line. CLMF
CC   synergistically induces, in the presence of interleukin-2, the cytolytic
CC   activity of lymphokine activated killer cells. CLMF is also capable of
CC   stimulating T-cell growth. The subunits may be used for preparing
CC   the CLMF protein and for preparing medicants for antitumour therapy.
XX
SQ   Sequence 1397 BP; 390 A; 307 C; 333 G; 367 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 18;  Length 1397;
  Best Local Similarity   87.0%;  Pred. No. 39;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     783 attctcggcaggtggaggtcagc 805


RESULT  19
AAT77849
ID   AAT77849 standard; DNA; 1397 BP.
XX
AC   AAT77849;
XX
DT   18-MAR-1998  (first entry)
XX
DE   40 kDa subunit of human cytotoxic lymphocyte maturation factor DNA.
XX
KW   Cytotoxic lymphocyte maturation factor; CLMF; 40 kDa subunit;
KW   interleukin-2; T-cell growth; antibody; lymphokine activated killer cell;
```

```
KW   CLMF bioactivity inhibition; affinity ligand; transplantation;
KW   cytotoxic T cell proliferation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1027
FT                   /*tag=  a
XX
PN   EP790255-A2.
XX
PD   20-AUG-1997.
XX
PF   09-DEC-1990;   90EP-0123670.
XX
PR   27-AUG-1990;   90US-0572284.
PR   22-DEC-1989;   89US-0455708.
PR   09-MAY-1990;   90US-0520935.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Chizzonite RA,  Gately MK,  Gubler UA,  Hulmes JD;
PI   Pan YE,  Podlanski FJ, Stern AS;
XX
DR   WPI; 1997-404698/38.
DR   P-PSDB; AAW24235.
XX
PT   Antibodies to cytotoxic lymphocyte maturation factor protein -
PT   useful as affinity ligands, assay reagents and medicaments
XX
PS   Disclosure; Fig 25; 80pp; English.
XX
CC   The present DNA sequence encodes the 40 kDa subunit of cytotoxic
CC   lymphocyte maturation factor (CLMF) protein. The natural CLMF protein is
CC   a 75 kDa heterodimer, where the 2 subunits (see AAT77850 for the 35 kDa
CC   subunit) are bonded together via one or more disulphide bonds. CLMF
CC   protein is produced by a human B lyphoblastoid cell line. CLMF
CC   synergistically induces, in the presence of interleukin-2, the cytolytic
CC   activity of lymphokine activated killer cells. CLMF is also capable of
CC   stimulating T-cell growth. A novel antibody directed against an epitope
CC   of CLMF protein (see AAW24237 for epitope) is capable of neutralising
CC   and/or inhibiting CLMF bioactivity. This antibody may be used as an
CC   affinity ligand for purifying the CLMF protein. The antibody can also
CC   be used as an assay reagent for detecting the CLMF protein, and as a
CC   medicament for selectively blocking the proliferation and activation of
CC   cytotoxic T cells, e.g. in transplantation.
XX
SQ   Sequence 1397 BP; 390 A; 307 C; 333 G; 367 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 18;  Length 1397;
  Best Local Similarity   87.0%;  Pred. No. 39;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    783 attctcggcaggtggaggtcagc 805


RESULT  20
AAT97969
ID   AAT97969 standard; DNA; 1397 BP.
XX
AC   AAT97969;
XX
DT   18-MAR-1998  (first entry)
XX
DE   40 kDa subunit of human cytotoxic lymphocyte maturation factor DNA.
XX
KW   Cytotoxic lymphocyte maturation factor; CLMF; 40 kDa subunit;
KW   interleukin-2; T-cell growth; antibody; lymphokine activated killer cell;
KW   CLMF bioactivity inhibition; affinity ligand; transplantation;
KW   cytotoxic T cell proliferation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1027
FT                   /*tag=  a
XX
PN   EP790308-A1.
XX
PD   20-AUG-1997.
XX
PF   09-DEC-1990;   90EP-0123670.
XX
PR   27-AUG-1990;   90US-0572284.
PR   22-DEC-1989;   89US-0455708.
PR   09-MAY-1990;   90US-0520935.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Chizzonite RA,  Gately MK,  Gubler UA,  Hulmes JD;
PI   Pan YE,  Podlanski FJ, Stern AS;
XX
DR   WPI; 1997-404733/38.
DR   P-PSDB; AAW36939.
XX
PT   Sub-unit of cytotoxic lymphocyte maturation factor protein - useful
PT   for producing protein and as antitumour agent
XX
PS   Disclosure; Fig 25; 83pp; English.
XX
CC   The present DNA sequence encodes a 40 kDa subunit of cytotoxic
CC   lymphocyte maturation factor (CLMF) protein. The natural CLMF protein is
CC   a 75 kDa heterodimer, where the 2 subunits (see AAT97970 for the 35 kDa
CC   subunit) are bonded together via one or more disulphide bonds. CLMF
CC   protein is produced by a human B lyphoblastoid cell line. CLMF
CC   synergistically induces, in the presence of interleukin-2, the cytolytic
CC   activity of lymphokine activated killer cells. CLMF is also capable of
CC   stimulating T-cell growth. The subunits may be used for preparing
CC   the CLMF protein and for preparing medicants for antitumour therapy.
XX
SQ   Sequence 1397 BP; 390 A; 307 C; 333 G; 367 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 18;  Length 1397;
  Best Local Similarity   87.0%;  Pred. No. 39;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    783 attctcggcaggtggaggtcagc 805


RESULT  21
AAV37767
ID   AAV37767 standard; cDNA; 1397 BP.
XX
AC   AAV37767;
XX
DT   17-SEP-1998  (first entry)
XX
DE   Cytotoxic lymphocyte maturation factor 40 kD subunit encoding cDNA.
XX
KW   Cytotoxic lymphocyte maturation factor; CLMF; cytokine; synergize;
KW   interleukin 2; IL-2; human; lymphoblastoid cell; detection; purifying;
KW   proliferation; cytotoxic T cell; transplantation; antibody; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1027
FT                   /*tag=  a
FT                   /product= "CLMF 40 kD subunit"
XX
```

PN US5780597-A.
XX
PD 14-JUL-1998.
XX
PF 02-JUN-1995; 95US-0460061.
XX
PR 24-MAR-1992; 92US-0857023.
PR 22-DEC-1989; 89US-0455708.
PR 09-MAY-1990; 90US-0520935.
PR 27-AUG-1990; 90US-0572284.
PR 02-MAR-1994; 94US-0205011.
PR 02-JUN-1995; 95US-0460061.
XX
PA (HOFF) HOFFMANN LA ROCHE INC.
XX
PI Chizzonite RA, Gately MK, Gubler UA, Hulmes JD;
PI Pan YE, Podlaski FJ, Stern AS;
XX
DR WPI; 1998-413150/35.
DR P-PSDB; AAW58725.
XX
PT New antibodies to cytotoxic lymphocyte maturation factor - useful
PT for detecting, purifying, and/or blocking proliferation and
PT activation of cytotoxic T cells, such as in transplantation(s)
XX
PS Example 10; Fig 25; 71pp; English.
XX
CC An isolated antibody has been developed which binds specifically to
CC cytotoxic lymphocyte maturation factor (CLMF) and neutralises CLMF
CC induced proliferation by more than 50% in a CLMF dependent T cell
CC growth assay and/or inhibits binding of at least 60% of the factor
CC to phytohaemagglutinin (PHA) activated peripheral blood lymphocyte
CC (PBL) blasts as determined in a CLMF receptor binding assay. CLMF is
CC a heterodimeric protein having a molecular weight band of 75 kD,
CC determined by sodium dodecyl sulphate-polyacrylamide gel electrophoresis
CC (SDS-PAGE) under non-reducing and/or reducing conditions, providing a
CC first subunit having a molecular weight band of 40 kD and a second
CC subunit having a molecular weight of 35 kD. The present sequence
CC encodes the CLMF 40 kD subunit, from an example of the present
CC invention. The antibody can be used for the purification and/or
CC detection of CLMF. It is also used in therapeutic treatments which
CC require selective blocking of proliferation and activation of
CC cytotoxic T cells (CTLs) such as in transplantation.
XX
SQ Sequence 1397 BP; 390 A; 309 C; 333 G; 365 T; 0 other;

Query Match 72.8%; Score 18.2; DB 19; Length 1397;
Best Local Similarity 87.0%; Pred. No. 39;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 783 attctcggcaggtggaggtcagc 805

RESULT 22
AAZ40021
ID AAZ40021 standard; DNA; 1399 BP.
XX
AC AAZ40021;
XX
DT 14-FEB-2000 (first entry)
XX
DE Human p40 coding sequence.
XX
KW Interleukin-12; IL-12; fusion protein; IL-12 p35 subunit; B7 protein;
KW IL-12 p40 subunit; gene therapy; tumour; leukaemia; ss.
XX
OS Homo sapiens.
XX
PN US5994104-A.
XX
PD 30-NOV-1999.
XX
PF 08-NOV-1996; 96US-0751767.
XX
PR 08-NOV-1996; 96US-0751767.
XX
PA (UNLO) ROYAL FREE HOSPITAL SCHOOL MED.
XX
PI Anderson RJ, Prentice HG, MacDonald ID;
XX
DR WPI; 2000-038261/03.
DR P-PSDB; AAY54919.
XX
PT Nucleic acid constructs encoding interleukin-12 fusion proteins useful
PT for treating leukemia and other cancers -
XX
PS Example; Fig 9; 73pp; English.
XX
CC This sequence encodes the human p40 protein sequence.
CC The invention relates to an isolated nucleic acid construct (I)
CC comprising a region encoding an interleukin-12 (IL-12) fusion protein
CC (comprising an IL-12 p35 subunit, an IL-12 p40 subunit and a linker
CC peptide (joining the subunits)) and a region encoding a B7 protein. (I)
CC may be used to produce IL-12 fusion proteins according to standard
CC recombinant DNA methodologies. The fusion proteins may be produced either
CC in vitro in a fermentation culture or in vivo as part of a gene therapy
CC protocol (in this case (I) is used to transform a patients cells, which
CC then secrete the functional polypeptide to supplement the patients own
CC production of IL-12 or to rectify mutations which lead to the expression
CC of inactive polypeptides). The fusion proteins produced in this way may
CC be used to treat any disease which responds to IL-12 such as tumours
CC (both solid and dispersed of the kidney, breast, colon, ovarian and
CC cervical tumours and melanomas) and in particular, tumours of the blood
CC such as leukaemia. Alternatively, the polypeptides may be used as
CC antigens in the production of antibodies to IL-12 and to assay for
CC agonists and antagonists of its activity. The antibodies and antagonists
CC may be used to inhibit the activity of IL-12. (I) may also be used
CC diagnostically as a probe which hybridizes to sequences encoding IL-12
CC and the antibodies may be used to detect the presence of IL-12
CC polypeptides in samples. They may be used diagnostically to quantitate
CC the expression of the polypeptide by patients and hence which subjects
CC may be in need of restorative therapy.
XX
SQ Sequence 1399 BP; 390 A; 310 C; 333 G; 366 T; 0 other;

Query Match 72.8%; Score 18.2; DB 21; Length 1399;
Best Local Similarity 87.0%; Pred. No. 39;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 785 attctcggcaggtggaggtcagc 807

RESULT 23
AAZ50892/c
ID AAZ50892 standard; cDNA; 2096 BP.
XX
AC AAZ50892;
XX
DT 31-MAY-2000 (first entry)
XX
DE Human receptor-associated protein cDNA from Incyte clone 3407686.
XX
KW Human receptor-associated protein; HRAP; Incyte clone 3407686;
KW cytostatic; immunomodulatory; antiinflammatory; cardiant; antianaemic;
KW antiarteriosclerotic; hepatotropic; antiarthritic antirheumatic;
KW antiasthmatic; osteopathic; antiallergic; antidiabetic; dermatological;
KW neuroprotective; diagnosis; treatment; prevention; reproductive disorder;
KW cardiovascular; cell proliferative; autoimmune; inflammatory; allergy;
KW gastrointestinal; atherosclerosis; cirrhosis; leukaemia; cancer; AIDS;
KW arthritis; anaemia; asthma; dermatitis; diabetes; osteoporosis;

```
KW   multiple sclerosis; irritable bowel syndrome; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             346..1449
FT                   /*tag= a
FT                   /product= "HRAP"
XX
PN   WO200008155-A2.
XX
PD   17-FEB-2000.
XX
PF   06-AUG-1999;    99WO-US17777.
XX
PR   07-AUG-1998;    98US-0160065.
PR   01-SEP-1998;    98US-0098703.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Hillman JL,  Yue H,  Lal P,  Tang YT,  Gorgone GA,  Guegler KJ;
PI   Corley NC,  Baughn MR;
XX
DR   WPI; 2000-205710/18.
DR   P-PSDB; AAY69990.
XX
PT   New human receptor-associated proteins (HRAP) useful for the diagnosis,
PT   treatment and prevention of cell proliferative, autoimmune,
PT   inflammatory, reproductive, cardiovascular, and gastrointestinal
PT   disorders -
XX
PS   Claim 9; Pages 91-92; 99pp; English.
XX
CC   The present sequence is a cDNA encoding human receptor-associated protein
CC   (HRAP) from Incyte clone 3407686 obtained from PROSTUS08 cDNA library.
CC   This sequence is expressed in reproductive, nervous and
CC   and cardiovascular tissues. HRAP has cytostatic, immunomodulatory,
CC   antiinflammatory, cardiant, antiarteriosclerotic, hepatotropic,
CC   antiarthritic, antirheumatic, osteopathic, antiallergic, antianaemic,
CC   antiasthmatic, antidiabetic, dermatological and neuroprotective
CC   activities. The present sequence is useful in the diagnosis, treatment
CC   and prevention of disorders associated with HRAP expression, especially
CC   cell proliferative, autoimmune/inflammatory, reproductive,
CC   cardiovascular and gastrointestinal disorders (e.g. atherosclerosis,
CC   cirrhosis, leukaemia, cancer, AIDS, arthritis, allergies, anaemia,
CC   asthma, dermatitis, diabetes, osteoporosis, multiple sclerosis and
CC   irritable bowel syndrome).
XX
SQ   Sequence 2096 BP; 422 A; 614 C; 572 G; 488 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 21;  Length 2096;
  Best Local Similarity   87.0%;  Pred. No. 41;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcag 24
          || ||||| ||||||| ||||||
Db     73 GAGCCTCGCCAGGTGCCGGTCAG 51


RESULT  24
AAA92406
ID   AAA92406 standard; cDNA; 2300 BP.
XX
AC   AAA92406;
XX
DT   15-JAN-2001  (first entry)
XX
DE   Human interleukin 2 encoding p35 and p40 proteins.
XX
KW   Human; interleukin 2; IL-2; p35; p40; immune response; tumour;
KW   infectious disease; bacteria; virus; parasite; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             19..678
FT                   /*tag= a
FT                   /product= "p35"
FT   CDS             1304..2290
FT                   /*tag= b
FT                   /product= "p40"
XX
PN   CN1258740-A.
XX
PD   05-JUL-2000.
XX
PF   31-DEC-1998;    98CN-0126748.
XX
PR   31-DEC-1998;    98CN-0126748.
XX
PA   (HUAC-) HUACHEN BIOLOGICAL TECHNOLOGY INST SHANG.
XX
PI   Cao X,  Zhang W,  Ju D;
XX
DR   WPI; 2000-558956/52.
DR   P-PSDB; AAB23795, AAB23796.
XX
PT   Recombinant adenovirus expressing human interleukin-12 and its
PT   preparation and use -
XX
PS   Example 3; Fig 4; 27pp; Chinese.
XX
CC   The present invention describes a recombinant adenovirus and its
CC   preparation and use, as well as medicine composite containing the
CC   recombinant adenovirus. The recombinant adenovirus is used clinically,
CC   has the advantages of low cost, easy use and high efficiency and is
CC   especially suitable for immune treatment of tumour and infectious
CC   diseases caused by bacteria, virus and parasite. The present sequence
CC   represents human interleukin 2 (IL-2) cDNA which encodes the p35 and
CC   p40 proteins, which is used in an example from the present invention.
XX
SQ   Sequence 2300 BP; 598 A; 591 C; 571 G; 540 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 21;  Length 2300;
  Best Local Similarity   87.0%;  Pred. No. 41;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db   2046 attctcggcaggtggaggtcagc 2068


RESULT  25
AAQ23586
ID   AAQ23586 standard; cDNA; 2362 BP.
XX
AC   AAQ23586;
XX
DT   16-SEP-1992  (first entry)
XX
DE   Natural killer cell stimulatory factor 40kD subunit.
XX
KW   NKSF; cytokine; gamma interferon; IFN; peripheral blood lymphocytes;
KW   IL-2; GM-CSF; granulocyte macrophage colony stimulating factor; PBL;
KW   ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             33..1019
FT                   /*tag= a
FT                   /product= pro 40 kD subunit of NKSF
XX
PN   WO9205256-A.
```

```
XX
PD   02-APR-1992.
XX
PF   04-SEP-1991;   91WO-US06332.
XX
PR   18-SEP-1990;   90US-0584941.
XX
PA   (GENE-) GENETICS INST INC.
PA   (WIST-) WISTAR INST.
XX
PI   Clark S C,  Hewick R,  Kobayashi M,  Perussia B,  Trinchieri G;
PI   Wong G G;
XX
DR   WPI; 1992-132124/16.
DR   P-PSDB; AAR22769.
XX
PT   New natural killer cell stimulatory factor - useful as an
PT   immunostimulant for inducing gamma-interferon and GM-CSF prodn.
PT   in human peripheral blood lymphocytes
XX
PS   Claim 12; Page 23; 79pp; English.
XX
CC   NKSF is capable of inducing the production of gamma interferon in
CC   human peripheral blood lymphocytes. It has subunits of 40 (AAR22769)
CC   and 30-35 kD (AAR23729). It is claimed that NKSF is useful in the
CC   treatment of bacterial and viral  infections e.g. AIDS.
CC   Degenerate probes were designed from tryptic fragments of NKSF
CC   and used to screen a cDNA library made from poly A RNA from PdBu
CC   induced 8866 cells to identify a cDNA encoding the small subunit
CC   of NKSF. cDNA prepared from 8866 poly A RNA was cloned into
CC   lambdaZAP and screened with a fragment within pNK-6 (A prior
CC   partial clone  containing the N-terminal non coding region of the
CC   40kD subunit). The library was then screened with probes designed
CC   from tryptic peptides as above,  to yield pNK162. This was ligated
CC   together with pNK-6 to form pNK40-4 which is believed to contain
CC   the full length cDNA for the  40 Kd NKSF subunit.
CC   See also AAR22769, AAQ24371, AAR23729.
XX
SQ   Sequence 2362 BP; 717 A; 482 C; 543 G; 620 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 13;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 41;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     775 attctcggcaggtggaggtcagc 797


RESULT  26
AAT48098
ID   AAT48098 standard; cDNA; 2362 BP.
XX
AC   AAT48098;
XX
DT   21-MAR-1997  (first entry)
XX
DE   Human interleukin-12 40 kD subunit gene.
XX
KW   Human; interleukin 12; IL-12; heterodimeric complex; immunogenic; ds;
KW   Leishmania; antigen; leishmaniasis; cell-mediated immunity; adjuvant;
KW   TH1 helper cell; TH2 helper cell; alum; cytokine; bacterium; cancer;
KW   vaccine; cell surface; membrane; glycoprotein; antigen presenting cell.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             33..1019
FT                   /*tag=  a
FT                   /product= interleukin-12 40 kD subunit
XX
PN   US5571515-A.
XX
PD   05-NOV-1996.
XX
PF   18-APR-1994;   94US-0229282.
XX
PR   17-JUN-1994;   94US-0265087.
PR   18-APR-1994;   94US-0229282.
XX
PA   (UYPE-) UNIV PENNSYLVANIA.
PA   (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI   Scott P,  Trinchieri G;
XX
DR   WPI; 1996-505347/50.
DR   P-PSDB; AAW07398.
XX
PT   Immunogenic compsn. to improve cell mediated immune response -
PT   contains soluble leishmania antigen and interleukin-12 as adjuvant
XX
PS   Disclosure; Column 17-22; 22pp; English.
XX
CC   This is the nucleotide sequence encoding the human interleukin (IL)-12
CC   40 kD subunit.  IL-12 is a heterodimeric complex composed of the 40 kD
CC   subunit and a 30 kD subunit (AAW07399).  The complex is used in a novel
CC   immunogenic composition comprising a soluble Leishmania antigen with
CC   IL-12, for protection against leishmaniasis.  The addition of IL-12
CC   improves cell-mediated immunity by inducing TH1 helper cells (as opposed
CC   to TH2 cells which are induced by alum adjuvant) and does not cause
CC   uncontrolled release of other cytokines (in contrast to bacterial
CC   adjuvants).  IL-12 can also be used as a cancer vaccine by association
CC   with the protein B7, a soluble, cell-surface (membrane)-bound
CC   glycoprotein which is expressed in antigen presenting cells.
XX
SQ   Sequence 2362 BP; 717 A; 483 C; 542 G; 620 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 17;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 41;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     775 attctcggcaggtggaggtcagc 797


RESULT  27
AAV42535
ID   AAV42535 standard; cDNA; 2362 BP.
XX
AC   AAV42535;
XX
DT   09-NOV-1998  (first entry)
XX
DE   Human interleukin-13 40 kDa subunit cDNA.
XX
KW   Interleukin-13; IL-13; human; cytokine; adjuvant; vaccine;
KW   tetanus toxoid; HIV-1; Mycobacterium tuberculosis; influenza A;
KW   hepatitis A; hepatitis B; hepatitis C; leprosy; leishmaniasis;
KW   malaria; AIDS; cancer; tumour; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             33..1019
FT                   /*tag=  a
XX
PN   WO9831384-A1.
XX
PD   23-JUL-1998.
XX
PF   22-JAN-1998;   98WO-US01500.
XX
PR   22-JAN-1997;   97US-0036342.
```

```
XX
PA   (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI   Montaner LJ;
XX
DR   WPI; 1998-413812/35.
DR   P-PSDB; AAW62992.
XX
PT   Enhancing immune response to antigen using interleukin-13 - useful
PT   in vaccine production, especially for antigens not responsive to
PT   known adjuvants e.g. Tetanus Toxoid and to treat e.g. HIV or cancer
XX
PS   Disclosure; Page 53-55; 95pp; English.
XX
CC   This cDNA sequence includes an open reading frame coding for the
CC   40 kDa subunit (see AAW62992) of human interleukin-13 (IL-13).  A
CC   claimed method for enhancing the immune response to a selected
CC   antigen in a host comprises administering an effective amount of
CC   IL-13 or a biologically active fragment ('IL-13'), simultaneously
CC   or sequentially with the antigen.  Also claimed are: (1) a method
CC   for enhancing the adjuvant effect of interleukin-12 (IL-12) on an
CC   antigen in a host by similarly administering IL-13 with the antigen
CC   and IL-12; (2) vaccine compositions comprising a selected antigen
CC   and IL-13, optionally with a second adjuvant; and (3) use of IL-13
CC   to prepare medicaments for administration as above.  The method is
CC   used to enhance the immune response of a host to a particular
CC   antigen, especially when the IL-13 is administered in medicaments as
CC   in (2), optionally further comprising IL-12 or functional fragments
CC   (especially in which the antigen is Tetanus Toxoid (TT), influenza
CC   A, HIV-1 p24, HIV-1 nef or Mycobacterium tuberculosis) (claimed).
CC   IL-13 was shown to be an effective adjuvant in both healthy or
CC   pathogen-infected mammals (e.g. HIV-infected humans) (claimed).
CC   The method and vaccines/medicaments of (1) and (2) may be used to
CC   treat viral infections (e.g. HIV, Influenza A, Hepatitis A, B or C
CC   etc.), bacterial infections (e.g. leprosy or tuberculosis),
CC   parasitic infections (e.g. leishmaniasis or malaria) and certain
CC   cancers and solid tumours.  Use of IL-13 may be preferable to use
CC   of other adjuvants since it is relatively stable in vivo (unlike
CC   IL-2), has a direct action on antigen presenting cells (unlike
CC   IL-12), and unlike bacterial adjuvants does not induce
CC   pro-inflammatory cytokines.  IL-13 can also adjuvant antigens not
CC   responsive to known adjuvants (e.g. TT).  IL-13 nucleic acids can
CC   also be used as adjuvants, operably linked to sequences encoding
CC   IL-2, IL-12 and/or the antigen.
XX
SQ   Sequence 2362 BP; 717 A; 483 C; 542 G; 620 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 19;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 41;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     775 attctcggcaggtggaggtcagc 797


RESULT  28
AAV15227
ID   AAV15227 standard; cDNA; 2362 BP.
XX
AC   AAV15227;
XX
DT   27-MAY-1998  (first entry)
XX
DE   Heterodimeric human interleukin 12 subunit 1 encoding cDNA.
XX
KW   Heterodimeric; human; interleukin 12; IL-12; vaccine; antigen;
KW   pathogen; immune response; microorganism; cancer; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             33..1019
FT                   /*tag=  a
FT                   /product= "heterodimeric human interleukin 12 subunit"
XX
PN   US5723127-A.
XX
PD   03-MAR-1998.
XX
PF   25-MAR-1996;   96US-0621493.
XX
PR   17-JUN-1994;   94US-0265087.
PR   18-APR-1994;   94US-0229282.
PR   25-MAR-1996;   96US-0621493.
XX
PA   (UYPE-) UNIV PENNSYLVANIA.
PA   (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI   Scott P,  Trinchieri G;
XX
DR   WPI; 1998-205996/18.
DR   P-PSDB; AAW44324.
XX
PT   Enhancing immune response to vaccine - comprises co-administering
PT   interleukin-12 with antigenic composition
XX
PS   Disclosure; Column 17-22; 23pp; English.
XX
CC   The present sequence encodes a heterodimeric human interleukin 12
CC   (IL-12) subunit (putatively the 40 kD subunit) shown in the present
CC   specification. The present invention describes a method for enhancing
CC   the immune response elicited by an antigenic composition. The method
CC   comprises administering IL-12 simultaneously or sequentially with the
CC   antigenic composition. The IL-12 can be used in therapeutic
CC   compositions, e.g. as an adjuvant in vaccines against pathogenic
CC   bacteria or viruses, especially in an amount of 0.1 mu g to 0.5 mg per
CC   25 mu g of antigen, or in cancer vaccines.
XX
SQ   Sequence 2362 BP; 717 A; 483 C; 542 G; 620 T; 0 other;


  Query Match             72.8%;  Score 18.2;  DB 19;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 41;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     775 attctcggcaggtggaggtcagc 797


RESULT  29
AAZ28862
ID   AAZ28862 standard; cDNA; 2362 BP.
XX
AC   AAZ28862;
XX
DT   01-FEB-2000  (first entry)
XX
DE   Human interleukin-12 40 kD subunit gene.
XX
KW   Human; interleukin-12; IL-12; antigen; cancer; immune response; ds.
XX
OS   Homo sapiens.
XX
PN   US5976539-A.
XX
PD   02-NOV-1999.
XX
PF   06-NOV-1997;   97US-0965688.
XX
PR   17-JUN-1994;   94US-0265087.
PR   25-MAR-1996;   96US-0621493.
PR   18-APR-1994;   94US-0229282.
XX
```

PA (UYPE-) UNIV PENNSYLVANIA.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI Trinchieri G, Scott P;
XX
DR WPI; 2000-021927/02.
DR P-PSDB; AAY44181.
XX
PT A new anticancer vaccine -
XX
PS Disclosure; Column 17-22; 23pp; English.
XX
CC This sequence represents the coding region for the human interleukin-12
CC (IL-12) 40 kD subunit (see AAZ28863 for the coding sequence of the 30 kD
CC subunit). IL-12 can be used in a composition with an antigen from a
CC cancer cell or a cancer cell transfected with an antigen, to stimulate
CC an immune response to the anticancer composition, treating cancer.
XX
SQ Sequence 2362 BP; 717 A; 483 C; 542 G; 620 T; 0 other;

Query Match 72.8%; Score 18.2; DB 21; Length 2362;
Best Local Similarity 87.0%; Pred. No. 41;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 775 attctcggcaggtggaggtcagc 797

RESULT 30
AAF56280
ID AAF56280 standard; DNA; 2362 BP.
XX
AC AAF56280;
XX
DT 18-APR-2001 (first entry)
XX
DE IL-12 vaccine adjuvant related DNA #1.
XX
KW Il-12; vaccine; adjuvant; immune; virus; ds.
XX
OS Unidentified.
XX
PN US6168923-B1.
XX
PD 02-JAN-2001.
XX
PF 01-MAR-1999; 99US-0260173.
XX
PR 17-JUN-1994; 94US-0265087.
PR 25-MAR-1996; 96US-0621493.
PR 06-NOV-1997; 97US-0965688.
PR 18-APR-1994; 94US-0229282.
XX
PA (UYPE-) UNIV PENNSYLVANIA.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI Trinchieri G, Scott P;
XX
DR WPI; 2001-122259/13.
XX
PT Use of interleukin 12 or a polynucleotide sequence encoding IL-12 as
PT adjuvant in vaccine compositions comprising an antigen from a pathogen
PT or a cancer cell -
XX
PS Disclosure; Column 17-22; 23pp; English.
XX
CC The present invention relates to a composition of an adjuvanting
CC amount of interleukin (IL)-12 and an antigen from a pathogenic
CC microorganism, designed to elicit a host cell mediated immune
CC response. The invention may be used to generate an immune response
CC against antigens from an intracellular parasite or a virus such as HIV,
CC hepatitis A, hepatitis B, hepatitis C, rabies virus, Herpes viruses,
CC papilloma virus, yellow fever virus, Epstein-Barr virus,
CC respiratory syncytical virus, parvovirus, chikungunya virus,
CC hemorrhagic fever viruses, and Klebsiella.
XX
SQ Sequence 2362 BP; 717 A; 483 C; 542 G; 620 T; 0 other;

Query Match 72.8%; Score 18.2; DB 22; Length 2362;
Best Local Similarity 87.0%; Pred. No. 41;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 775 attctcggcaggtggaggtcagc 797

RESULT 31
AAZ40415
ID AAZ40415 standard; DNA; 5686 BP.
XX
AC AAZ40415;
XX
DT 15-FEB-2000 (first entry)
XX
DE Plasmid pIN1143 containing human IL-12 sequence.
XX
KW Wild type; human; interferon-alpha; plasmid; cytomegalovirus; CMV;
KW promoter; growth hormone; untranslated region; UTR; mammal; disease;
KW cancer; intron; ss.
XX
OS Synthetic.
XX
PN WO9947678-A2.
XX
PD 23-SEP-1999.
XX
PF 12-MAR-1999; 99WO-US05394.
XX
PR 19-MAR-1998; 98US-0078654.
XX
PA (GENE-) GENEMEDICINE INC.
XX
PI Nordstrom J, Pericle F, Rolland A, Ralston R;
XX
DR WPI; 1999-562116/47.
XX
PT New plasmids containing an interferon-alpha coding sequence, used for
PT the treatment of a mammalian condition or disease, particularly cancer
PT -
XX
PS Disclosure; Fig 2; 137pp; English.
XX
CC The invention relates to a novel plasmid comprising a cytomegalovirus
CC (CMV) promoter transcriptionally linked with an interferon alpha
CC (IFN-alpha) coding sequence, and a growth hormone 3'-untranslated
CC region (UTR). This sequence represents the plasmid pIN1143 which
CC contains the human interleukin 12 (IL-12) gene. The plasmids can be
CC used for treating a mammalian condition or disease, particularly cancer.
XX
SQ Sequence 5686 BP; 1367 A; 1517 C; 1446 G; 1356 T; 0 other;

Query Match 72.8%; Score 18.2; DB 20; Length 5686;
Best Local Similarity 87.0%; Pred. No. 44;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 atcctcgtcaggtgcaggtcagc 25
|| |||| |||||| ||||||||
Db 3961 attctcggcaggtggaggtcagc 3983

RESULT 32

```
AAZ40023
ID   AAZ40023 standard; DNA; 6139 BP.
XX
AC   AAZ40023;
XX
DT   14-FEB-2000  (first entry)
XX
DE   Interleukin-12 fusion protein, Flexi-12, coding sequence.
XX
KW   Interleukin-12; IL-12; fusion protein; IL-12 p35 subunit; B7 protein;
KW   IL-12 p40 subunit; gene therapy; tumour; leukaemia; Flexi-12; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   US5994104-A.
XX
PD   30-NOV-1999.
XX
PF   08-NOV-1996;   96US-0751767.
XX
PR   08-NOV-1996;   96US-0751767.
XX
PA   (UNLO ) ROYAL FREE HOSPITAL SCHOOL MED.
XX
PI   Anderson RJ,  Prentice HG,  MacDonald ID;
XX
DR   WPI; 2000-038261/03.
DR   P-PSDB; AAY54921.
XX
PT   Nucleic acid constructs encoding interleukin-12 fusion proteins useful
PT   for treating leukemia and other cancers -
XX
PS   Example; Fig 11; 73pp; English.
XX
CC   This sequence encodes the human interleukin-12 fusion protein sequence,
CC   designated Flexi-12. The invention relates to an isolated nucleic acid
CC   construct (I) comprising a region encoding an interleukin-12 (IL-12)
CC   fusion protein (comprising an IL-12 p35 subunit, an IL-12 p40 subunit and
CC   a linker peptide (joining the subunits)) and a region encoding a B7
CC   protein. (I) may be used to produce IL-12 fusion proteins according to
CC   standard recombinant DNA methodologies. The fusion proteins may be
CC   produced either in vitro in a fermentation culture or in vivo as part of
CC   a gene therapy protocol (in this case (I) is used to transform a patients
CC   cells, which then secrete the functional polypeptide to supplement the
CC   patients own production of IL-12 or to rectify mutations which lead to
CC   the expression of inactive polypeptides). The fusion proteins produced in
CC   this way may be used to treat any disease which responds to IL-12 such as
CC   tumours (both solid and dispersed of the kidney, breast, colon, ovarian
CC   and cervical tumours and melanomas) and in particular, tumours of the
CC   blood such as leukaemia. Alternatively, the polypeptides may be used as
CC   antigens in the production of antibodies to IL-12 and to assay for
CC   agonists and antagonists of its activity. The antibodies and antagonists
CC   may be used to inhibit the activity of IL-12. (I) may also be used
CC   diagnostically as a probe which hybridizes to sequences encoding IL-12
CC   and the antibodies may be used to detect the presence of IL-12
CC   polypeptides in samples. They may be used diagnostically to quantitate
CC   the expression of the polypeptide by patients and hence which subjects
CC   may be in need of restorative therapy.
XX
SQ   Sequence 6139 BP; 1552 A; 1538 C; 1558 G; 1491 T; 0 other;


  Query Match            72.8%;  Score 18.2;  DB 21;  Length 6139;
  Best Local Similarity   87.0%;  Pred. No. 45;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db    3964 attctcggcaggtggaggtcagc 3986


RESULT  33
```

```
AAF98655
ID   AAF98655 standard; DNA; 316 BP.
XX
AC   AAF98655;
XX
DT   02-JUL-2001  (first entry)
XX
DE   Human ovarian cancer cell expressed sequence 8094.
XX
KW   Human; ovarian cancer; identification; detection; characterisation;
KW   tumour; kinase; marker; cytostatic; antisense gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200118542-A2.
XX
PD   15-MAR-2001.
XX
PF   01-SEP-2000; 2000WO-US24199.
XX
PR   03-SEP-1999;   99US-0152547.
PR   16-MAR-2000; 2000US-0190347.
PR   21-MAR-2000; 2000US-0191321.
PR   31-MAY-2000; 2000US-0208382.
PR   20-JUL-2000; 2000US-0220467.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lee J,  Thompsho P,  Lillie J;
XX
DR   WPI; 2001-211428/21.
XX
PT   Detection, assessment, prevention and therapy of ovarian cancer,
PT   comprises detecting changes in the expression of a variety of markers -
XX
PS   Claim 1; Page 993; 1198pp; English.
XX
CC   The present invention describes a method for assessing whether a patient
CC   is afflicted with ovarian cancer by comparing: (1) the expression of a
CC   marker (I) (see AAF98594 to AAF98730), in a patient sample; and (2) the
CC   normal level of expression of (I) in a control non-ovarian cancer
CC   sample, where a significant difference between the level of expression
CC   in (a) and (b) is an indication that the patient is afflicted with
CC   ovarian cancer. (I) have cytostatic activities and can be used in
CC   antisense gene therapy. The method, compositions and kits from the
CC   present invention can be used for: (1) assessing and treating ovarian
CC   cancer; (2) making isolated hybridoma, which produces an antibody useful
CC   for ovarian cancer assessment; and (3) inhibiting ovarian cancer in a
CC   patient. AAF98573 to AAF98593 represent human kinase marker primers and
CC   probes which are used in the exemplification of the present invention.
XX
SQ   Sequence 316 BP; 66 A; 93 C; 82 G; 73 T; 2 other;


  Query Match            71.2%;  Score 17.8;  DB 22;  Length 316;
  Best Local Similarity   86.4%;  Pred. No. 52;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtc 22
           | |||| |||||||||| |||
Db     264 ccatccgcgtcaggtgcangtc 285


RESULT  34
AAC75563/c
ID   AAC75563 standard; cDNA; 586 BP.
XX
AC   AAC75563;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF1118 polynucleotide sequence SEQ ID NO:2235.
XX
```

```
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB41354.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 1649; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 586 BP; 122 A; 185 C; 156 G; 123 T; 0 other;


  Query Match             68.8%;  Score 17.2;  DB 21;  Length 586;
  Best Local Similarity   86.4%;  Pred. No. 1e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcag 24
           |||||||||||||||| ||  ||
Db     304 ATCCTCGTCAGGTGCCGGGTAG 283
```

```
RESULT  35
AAH43851/c
ID   AAH43851 standard; cDNA; 6372 BP.
XX
AC   AAH43851;
XX
DT   04-SEP-2001  (first entry)
XX
DE   Human CLASP-7 encoding cDNA sequence SEQ ID NO:1.
XX
KW   Human; CLASP-7; cadherin-like asymmetry protein; immune response;
KW   neuroprotective; antidiabetic; immunosuppressive; antirheumatic;
KW   antiarthritic; hypotensive; anti-HIV; cytostatic; immunostimulant;
KW   antianaemic; antiinflammatory; ophthalmological; nephrotrophic;
KW   antithyroid; antiasthmatic; antiallergic; antibacterial; gene therapy;
KW   chromosome 19q13.2; autoimmune disease; multiple sclerosis; toxaemia;
KW   juvenile diabetes; rheumatoid arthritis; pruritic urticarcial papule;
KW   hypertension; Rh incompatibility; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             13..6156
FT                   /*tag= a
FT                   /product= "CLASP-7"
XX
PN   WO200142295-A2.
XX
PD   14-JUN-2001.
XX
PF   13-DEC-2000; 2000WO-US34152.
XX
PR   13-DEC-1999;   99US-0170453.
PR   14-JAN-2000; 2000US-0176195.
PR   14-FEB-2000; 2000US-0182296.
PR   11-APR-2000; 2000US-0196267.
PR   11-APR-2000; 2000US-0196460.
PR   11-APR-2000; 2000US-0196527.
PR   11-APR-2000; 2000US-0196528.
PR   11-APR-2000; 2000US-0547276.
PR   13-OCT-2000; 2000US-0240503.
PR   13-OCT-2000; 2000US-0240508.
PR   13-OCT-2000; 2000US-0240539.
PR   13-OCT-2000; 2000US-0240543.
XX
PA   (ARBO-) ARBOR VITA CORP.
XX
PI   Lu P,  Garman JD,  Candia AF;
XX
DR   WPI; 2001-381641/40.
DR   P-PSDB; AAB99541.
XX
PT   Novel cadherin-like asymmetry protein-7 and polynucleotides encoding
PT   the polypeptide, useful for treating autoimmune disease,
PT   hypersensitivity, preventing transplant rejection by modulating immune
PT   response -
XX
PS   Claim 1; Fig 5; 151pp; English.
XX
CC   The present invention describes a human cadherin-like asymmetry protein
CC   (CLASP), designated CLASP-7. The CLASP-7 protein (I) and its encoding
CC   nucleotide sequence (II) have activities including: neuroprotective;
CC   antidiabetic; immunosuppressive; antirheumatic; antiarthritic; anti-HIV;
CC   hypotensive; cytostatic; immunostimulant; antianaemic; antiinflammatory;
CC   ophthalmological; nephrotrophic; antithyroid; antiasthmatic;
CC   antiallergic; and antibacterial. (II) and CLASP-7 antibodies (III) are
CC   useful for detecting the CLASP-7 polypeptide. (II) is useful for
CC   producing (I) by recombinant methods. (I) or its fragment are useful for
CC   inhibiting an immune response in a cell such as T cell or B cell. A
CC   pharmaceutical composition (C), comprising (I) or (II), can be useful
CC   for treating CLASP-7-mediated disease such as an autoimmune disease
```

```
CC   caused or exacerbated by increased activity of T helper cells. Autoimmune
CC   diseases which can be treated using (C) include multiple sclerosis,
CC   juvenile diabetes and rheumatoid arthritis. (I) is useful for treating
CC   toxaemia or pregnancy induced hypertension, pruritic urticarcial papules
CC   and Rh incompatibility. (I) is also useful as a diagnostic reagent for
CC   immune and other disorders, since diseases characterised by
CC   overproduction or depletion of lymphocytes in blood or other organs may
CC   be detected by monitoring the level of (I) or its mRNA. CLASP-7 has been
CC   mapped to the chromosomal location 19q13.2. The present sequence
CC   encodes the human CLASP-7 protein as given in the present invention.
XX
SQ   Sequence 6372 BP; 1273 A; 2050 C; 1849 G; 1200 T; 0 other;


  Query Match             68.8%;  Score 17.2;  DB 22;  Length 6372;
  Best Local Similarity   86.4%;  Pred. No. 1.3e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcag 24
           ||||||||||||||| ||  ||
Db     735 ATCCTCGTCAGGTGCCGGGTAG 714


RESULT  36
AAC76615/c
ID   AAC76615 standard; cDNA; 5269 BP.
XX
AC   AAC76615;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF2170 polynucleotide sequence SEQ ID NO:4339.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42406.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 3533-3536; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 5269 BP; 1031 A; 1619 C; 1568 G; 1049 T; 2 other;


  Query Match             67.2%;  Score 16.8;  DB 21;  Length 5269;
  Best Local Similarity   90.0%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcagg 20
           |||||| ||||||| |||||
Db     439 CGATCCACGTCAGGCGCAGG 420


RESULT  37
AAQ75832
ID   AAQ75832 standard; DNA; 34 BP.
XX
AC   AAQ75832;
XX
DT   18-AUG-1995  (first entry)
XX
DE   Sense primer to amplify Non-A Non-B hepatitis virus for analysis.
XX
KW   Non-A Non-B hepatitis virus; structural region; cDNA to genomic RNA;
KW   detection; reagent; anti-Non-A Non-B hepatitis virus antibody;
KW   vaccine; antigen; epitope; diagnosis; ss.
XX
OS   Synthetic.
XX
PN   EP628572-A.
XX
PD   14-DEC-1994.
XX
PF   27-MAY-1994;   94EP-0108256.
XX
PR   28-MAY-1993;   93JP-0126709.
PR   02-MAR-1994;   94JP-0032201.
XX
PA   (ARIM/) ARIMA T.
PA   (EISA ) EISAI CO LTD.
XX
PI   Aoyama M,  Arima T,  Hosoda T,  Iwasaki Y,  Obara T;
PI   Sawada T, Tohmatsu J;
XX
DR   WPI; 1995-015655/03.
XX
PT   New non-A non-B hepatitis virus sub-type - used to develop prods.
PT   for detection, diagnosis, prevention and treatment of non-A non-B
PT   hepatitis.
XX
PS   Example 2; Page 49; 59pp; English.
```

```
XX
CC   This primer is based on nucleotides 778-800 of AAQ75818 (a part of
CC   the Non-A Non-B hepatitis virus genome encoding the non-structural
CC   protein). It is used in conjunction with AAQ75833 to amplify nucleotides
CC   800-1326 of AAQ75818.  The nucleotide sequences (see also AAQ75817-19)
CC   were isolated from the plasma of donors in Japan with high s-GTP levels,
CC   and were found to be different from previously reported NANB hepatitis
CC   viruses. The DNA can be used as a reagent for detecting the NANB
CC   hepatitis viral gene. The polypeptides can be used as reagents for
CC   detecting anti-NANB hepatitis antibodies or as a NANB hepatitis viral
CC   vaccine.
XX
SQ   Sequence 34 BP; 6 A; 9 C; 12 G; 7 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 16;  Length 34;
  Best Local Similarity   82.6%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           |||||||||||||||| | | | |
Db       8 gatcctcgtcaggtgaaaggccg 30


RESULT  38
AAF28005
ID   AAF28005 standard; cDNA; 220 BP.
XX
AC   AAF28005;
XX
DT   08-MAY-2001  (first entry)
XX
DE   Human TR13 coding sequence fragment SEQ ID NO: 14.
XX
KW   Human; tumour necrosis factor receptor; TR13; TR14; infection;
KW   cancer; autoimmune disease; allergy; inflammatory disease;
KW   graft rejection; apoptosis; cardiovascular disease; aneurysm; ss.
XX
OS   Homo sapiens.
XX
PN   WO200105834-A1.
XX
PD   25-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-US19343.
XX
PR   16-JUL-1999;   99US-0144087.
PR   18-AUG-1999;   99US-0149450.
PR   20-AUG-1999;   99US-0149712.
PR   10-SEP-1999;   99US-0153089.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Young PE;
XX
DR   WPI; 2001-112682/12.
XX
PT   Nucleic acids encoding 2 human tumor necrosis factor receptor
PT   polypeptides ((TR13) and (TR14)), useful for the prevention, diagnosis
PT   and treatment of, e.g. cancers, acquired immune deficiency syndrome and
PT   hypohidrotic ectodermal dysplasia -
XX
PS   Claim 31; Page 382; 418pp; English.
XX
CC   The present invention provides the protein and coding sequences of the
CC   human tumour necrosis factor receptors TR13 and TR14. These sequences are
CC   useful in the diagnosis and treatment of many diseases, including cancer,
CC   autoimmune diseases, cardiovascular disorders, allergies,
CC   neurodegenerative diseases, graft rejection, inflammation, aneurysms and
CC   infections.
XX
SQ   Sequence 220 BP; 43 A; 63 C; 57 G; 56 T; 1 other;


  Query Match             66.4%;  Score 16.6;  DB 22;  Length 220;
  Best Local Similarity   82.6%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtca 23
           | |||| ||||||||||||  ||
Db      59 ccatccgcgtcaggtgcagtcca 81


RESULT  39
AAF28044
ID   AAF28044 standard; cDNA; 220 BP.
XX
AC   AAF28044;
XX
DT   08-MAY-2001  (first entry)
XX
DE   Human TR13 coding sequence fragment SEQ ID NO: 55.
XX
KW   Human; tumour necrosis factor receptor; TR13; TR14; infection;
KW   cancer; autoimmune disease; allergy; inflammatory disease;
KW   graft rejection; apoptosis; cardiovascular disease; aneurysm; ss.
XX
OS   Homo sapiens.
XX
PN   WO200105834-A1.
XX
PD   25-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-US19343.
XX
PR   16-JUL-1999;   99US-0144087.
PR   18-AUG-1999;   99US-0149450.
PR   20-AUG-1999;   99US-0149712.
PR   10-SEP-1999;   99US-0153089.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Ni J,  Young PE;
XX
DR   WPI; 2001-112682/12.
XX
PT   Nucleic acids encoding 2 human tumor necrosis factor receptor
PT   polypeptides ((TR13) and (TR14)), useful for the prevention, diagnosis
PT   and treatment of, e.g. cancers, acquired immune deficiency syndrome and
PT   hypohidrotic ectodermal dysplasia -
XX
PS   Claim 33; Page 407; 418pp; English.
XX
CC   The present invention provides the protein and coding sequences of the
CC   human tumour necrosis factor receptors TR13 and TR14. These sequences are
CC   useful in the diagnosis and treatment of many diseases, including cancer,
CC   autoimmune diseases, cardiovascular disorders, allergies,
CC   neurodegenerative diseases, graft rejection, inflammation, aneurysms and
CC   infections.
XX
SQ   Sequence 220 BP; 43 A; 63 C; 57 G; 56 T; 1 other;


  Query Match             66.4%;  Score 16.6;  DB 22;  Length 220;
  Best Local Similarity   82.6%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtca 23
           | |||| ||||||||||||  ||
Db      59 ccatccgcgtcaggtgcagtcca 81


RESULT  40
AAC98199
ID   AAC98199 standard; cDNA; 625 BP.
XX
```

```
AC   AAC98199;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human colon cancer antigen nucleotide sequence SEQ ID NO:209.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW   immunomodulatory; muscular; gynaecological; gastrointestinal;
KW   nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW   neural disorder; immune system disorder; muscular disorder;
KW   reproductive disorder; gastrointestinal disorder; renal disorder;
KW   infectious disease; cardiovascular disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055351-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05883.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587534/55.
DR   P-PSDB; AAB53442.
XX
PT   Colon cancer associated gene sequences, referred to as colon cancer
PT   antigens, useful for the treatment, prevention, and diagnosis of colon
PT   disorders such as colon cancer -
XX
PS   Claim 1; Page 633; 2104pp; English.
XX
CC   AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC   called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC   human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC   neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC   vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC   can be used in gene therapy. The colon cancer antigen polynucleotides,
CC   proteins and antibodies to the proteins are useful for the prevention,
CC   treatment and diagnosis of colon disorders, such as colon cancer. The
CC   polynucleotides may be used in diagnostics and research, such as for
CC   chromosome identification, and as hybridisation probes. The proteins
CC   may also be used to prevent diseases such as neural disorders, immune
CC   system disorders, muscular disorders, reproductive disorders,
CC   gastrointestinal disorders, wounds, renal disorders, infectious
CC   diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC   AAB54007 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 625 BP; 139 A; 177 C; 143 G; 156 T; 10 other;

  Query Match             66.4%;  Score 16.6;  DB 21;  Length 625;
  Best Local Similarity   82.6%;  Pred. No. 1.9e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtca 23
           | ||||  ||||||||||||  ||
Db     464 ccatccgcgtcaggtgcagtcca 486


RESULT  41
AAF14300
ID   AAF14300 standard; cDNA; 651 BP.
XX
AC   AAF14300;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Aspergillus oryzae EST SEQ ID NO:6823.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS   Aspergillus oryzae.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999;   99US-0273623.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
PA   (NOVO ) NOVO NORDISK AS.
XX
PI   Berka RM,  Rey MW,  Shuster JR,  Kauppinen S,  Clausen IG,  Olsen PB;
XX
DR   WPI; 2000-594572/56.
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
PS   Claim 88; Page 2778; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 651 BP; 185 A; 171 C; 162 G; 133 T; 0 other;

  Query Match             66.4%;  Score 16.6;  DB 21;  Length 651;
  Best Local Similarity   82.6%;  Pred. No. 1.9e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtca 23
           || ||||||||| | |||||  |
Db     552 cggtcctcgtcagttccaggtga 574


RESULT  42
AAF71312
ID   AAF71312 standard; DNA; 681 BP.
XX
```

AC AAF71312;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum HA protein nucleotide sequence SEQ ID NO:349.
XX
KW Corynebacterium glutamicum; homeostasis; adaptation; HA protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; genetic engineering;
KW Brevibacterium; environmental condition; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100842-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00911.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031636.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032127.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032129.
PR 09-JUL-1999; 99DE-1032226.
PR 14-JUL-1999; 99DE-1032920.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1032930.
PR 14-JUL-1999; 99DE-1032933.
PR 14-JUL-1999; 99DE-1032935.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033002.
PR 14-JUL-1999; 99DE-1033003.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041390.
PR 31-AUG-1999; 99DE-1041391.
PR 03-SEP-1999; 99DE-1042088.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-061974/07.
DR P-PSDB; AAB79197.
XX
PT New isolated Corynebacterium glutamicum nucleic acid for production or
PT modulation of production of fine chemicals such as amino acids,
PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins
PT or enzymes -
XX
PS Claim 3; Page 597-598; 712pp; English.
XX
CC AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis
CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The
CC C. glutamicum HA genes (I) can be used in vectors for expression in host
CC cells and production of fine chemicals, such as, an organic acid,
CC proteinogenic or nonproteinogenic amino acid (preferred), purine or
CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated
CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,
CC polyketide or enzyme. The amino acids produced can be lysine, glutamine,
CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,
CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,
CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can
CC be modulated. The presence of (I) or HA proteins encoded by then are
CC used for diagnosing the presence or activity of Corynebacterium
CC diphtheriae. (I) can be used to map the C. glutamicum genome or can be
CC used as markers for genetically engineered Corynebacterium or
CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain
CC homeostasis in C. glutamicum or help the microorganism to adapt to
CC different environmental conditions.
XX
SQ Sequence 681 BP; 149 A; 164 C; 221 G; 147 T; 0 other;

Query Match 66.4%; Score 16.6; DB 22; Length 681;
Best Local Similarity 82.6%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 gatcctcgtcaggtgcaggtcag 24
|||| ||| ||| ||||||| ||
Db 78 gatcgtcgccagctgcaggtgag 100

RESULT 43
AAX18177
ID AAX18177 standard; cDNA; 924 BP.
XX
AC AAX18177;
XX
DT 07-MAY-1999 (first entry)
XX
DE Canine mature CLAF p40 subunit coding sequence.
XX
KW CLAF; canine lymphocyte activation factor; p35 subunit; p40 subunit; dog;
KW viral disease; canine distemper; canine parvovirus; autoimmune disease;
KW canine infectious hepatitis; ds.
XX
OS Canis sp.
XX
PN WO9855511-A1.
XX
PD 10-DEC-1998.
XX
PF 26-MAY-1998; 98WO-JP02295.
XX
PR 03-JUN-1997; 97JP-0161936.
XX
PA (KAGA) CHEMO-SERO-THERAPEUTIC RES INST.
XX
PI Eda Y, Imamura T, Maeda H, Tokiyoshi S;
XX
DR WPI; 1999-070260/06.
DR P-PSDB; AAW74323.
XX
PT Canine cytokine protein which activates canine cytotoxic T
PT lymphocytes - is useful for treatment of viral and autoimmune
PT diseases in dogs
XX
PS Claim 8; Page 32; 51pp; Japanese.
XX
CC This sequence encodes the canine lymphocyte activation factor (CLAF) of
CC the invention. CLAF activates lymphocytes, especially canine cytotoxic
CC T cells. The protein contains a p35 and a p40 subunit. CLAF or antibodies
CC recognising it may be used in the treatment of viral diseases in dogs
CC (such as canine distemper, canine parvovirus and canine infectious
CC hepatitis). The antibodies may also be used in isolation of CLAF from
CC culture of the transformant host cells by affinity chromatography. The
CC p40 homodimer can be used for the treatment of autoimmune diseases in
CC dogs.
XX
SQ Sequence 924 BP; 273 A; 222 C; 234 G; 195 T; 0 other;

Query Match 66.4%; Score 16.6; DB 20; Length 924;
Best Local Similarity 82.6%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| || ||| ||||||||
Db     683 attctcggcacgtggaggtcagc 705


RESULT  44
AAV13816
ID   AAV13816 standard; cDNA to mRNA; 990 BP.
XX
AC   AAV13816;
XX
DT   14-MAY-1998  (first entry)
XX
DE   Canine IL-12 P40 subunit cDNA.
XX
KW   Canine; interleukin-12 P40 subunit; IL-12 P40 subunit; antitumour;
KW   antiviral; vaccine adjuvant; ds.
XX
OS   Canis sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..990
FT                   /*tag=  a
FT                   /product= interleukin-12_P40_subunit
XX
PN   JP10036397-A.
XX
PD   10-FEB-1998.
XX
PF   08-NOV-1996;   96JP-0296789.
XX
PR   23-MAY-1996;   96JP-0128104.
PR   08-NOV-1995;   95JP-0289729.
XX
PA   (TORA ) TORAY IND INC.
XX
DR   WPI; 1998-174914/16.
DR   P-PSDB; AAW41790.
XX
PT   Canine interleukin 12 - comprises P40 and P35 subunits; useful in
PT   veterinary medicine, e.g. antitumour, antivirus and vaccine adjuvant
PT   activities are expected
XX
PS   Claim 8; Page 8; 12pp; Japanese.
XX
CC   The present sequence encodes a canine interleukin-12 (IL-12) P40
CC   subunit. A canine IL-12 comprising a P40 and P35 subunit is capable
CC   of inducing an antiviral activating factor and the expression of
CC   class II MHC molecules in canine tumour cells, stimulating
CC   proliferation of canine blastogenic lymphocytes and activating
CC   canine leukocytes to inhibit canine tumour cells. The canine IL-12
CC   can be used in veterinary medicines, e.g. antitumour, antivirus and
CC   vaccine adjuvant activities are expected.
XX
SQ   Sequence 990 BP; 278 A; 243 C; 251 G; 218 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 19;  Length 990;
  Best Local Similarity   82.6%;  Pred. No. 2e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| || ||| ||||||||
Db     746 attctcggcacgtggaggtcagc 768


RESULT  45
AAV13817
ID   AAV13817 standard; cDNA to mRNA; 990 BP.
XX
AC   AAV13817;
XX
DT   14-MAY-1998  (first entry)
XX
DE   Canine IL-12 P40 subunit cDNA.
XX
KW   Canine; interleukin-12 P40 subunit; IL-12 P40 subunit; antitumour;
KW   antiviral; vaccine adjuvant; ds.
XX
OS   Canis sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..990
FT                   /*tag=  a
FT                   /product= interleukin-12_P40_subunit
XX
PN   JP10036397-A.
XX
PD   10-FEB-1998.
XX
PF   08-NOV-1996;   96JP-0296789.
XX
PR   23-MAY-1996;   96JP-0128104.
PR   08-NOV-1995;   95JP-0289729.
XX
PA   (TORA ) TORAY IND INC.
XX
DR   WPI; 1998-174914/16.
DR   P-PSDB; AAW41791.
XX
PT   Canine interleukin 12 - comprises P40 and P35 subunits; useful in
PT   veterinary medicine, e.g. antitumour, antivirus and vaccine adjuvant
PT   activities are expected
XX
PS   Claim 10; Pages 10-11; 12pp; Japanese.
XX
CC   The present sequence encodes a canine interleukin-12 (IL-12) P40
CC   subunit. A canine IL-12 comprising a P40 and P35 subunit is capable
CC   of inducing an antiviral activating factor and the expression of
CC   class II MHC molecules in canine tumour cells, stimulating
CC   proliferation of canine blastogenic lymphocytes and activating
CC   canine leukocytes to inhibit canine tumour cells. The canine IL-12
CC   can be used in veterinary medicines, e.g. antitumour, antivirus and
CC   vaccine adjuvant activities are expected.
XX
SQ   Sequence 990 BP; 279 A; 244 C; 249 G; 218 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 19;  Length 990;
  Best Local Similarity   82.6%;  Pred. No. 2e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| || ||| ||||||||
Db     746 attctcggcacgtggaggtcagc 768


Search completed: December 26, 2001, 12:19:37
Job time: 6938 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:27:10 ; Search time 141.92 Seconds
(without alignments)
39.895 Million cell updates/sec

Title: US-09-396-196F-9
Perfect score: 25
Sequence: 1 cgatcctcgtcaggtgcaggtcagc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
	2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
	3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
	4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
	5	18.2	72.8	987	1	US-08-186-529-1	Sequence 1, Appli
	6	18.2	72.8	987	1	US-08-640-386A-1	Sequence 1, Appli
	7	18.2	72.8	987	4	US-08-848-760B-24	Sequence 24, Appl
	8	18.2	72.8	1018	2	US-08-184-009-194	Sequence 194, App
	9	18.2	72.8	1018	2	US-08-458-356-194	Sequence 194, App
	10	18.2	72.8	1018	4	US-08-460-736-194	Sequence 194, App
	11	18.2	72.8	1399	2	US-08-751-767A-3	Sequence 3, Appli
	12	18.2	72.8	1560	2	US-08-751-767A-11	Sequence 11, Appl
	13	18.2	72.8	1623	2	US-08-751-767A-9	Sequence 9, Appli
	14	18.2	72.8	2362	1	US-08-265-087-1	Sequence 1, Appli
	15	18.2	72.8	2362	1	US-08-621-493-1	Sequence 1, Appli
	16	18.2	72.8	2362	2	US-08-965-688-1	Sequence 1, Appli
	17	18.2	72.8	2362	4	US-09-260-173-1	Sequence 1, Appli
	18	18.2	72.8	6139	2	US-08-751-767A-7	Sequence 7, Appli
	19	16.6	66.4	990	4	US-09-079-984A-1	Sequence 1, Appli
	20	16.6	66.4	990	4	US-09-079-984A-11	Sequence 11, Appl
c	21	16.6	66.4	3449	3	US-09-049-672A-16	Sequence 16, Appl
c	22	16.2	64.8	3509	1	US-08-175-471-6	Sequence 6, Appli
c	23	16.2	64.8	3509	2	US-08-429-054A-12	Sequence 12, Appl
c	24	16.2	64.8	3509	2	US-08-718-777-6	Sequence 6, Appli
c	25	16.2	64.8	3509	3	US-09-078-862-2	Sequence 2, Appli
c	26	16.2	64.8	3509	3	US-09-051-341-6	Sequence 6, Appli
c	27	15.8	63.2	2095	1	US-08-336-408B-3	Sequence 3, Appli
c	28	15.8	63.2	2095	5	PCT-US91-00399-3	Sequence 3, Appli
c	29	15.6	62.4	3336	6	5457037-1	Patent No. 5457037
	30	15.6	62.4	80246	4	US-09-078-294-4	Sequence 4, Appli
	31	15.6	62.4	80595	4	US-09-078-294-3	Sequence 3, Appli
	32	15.6	62.4	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	33	15.6	62.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	34	15.4	61.6	1722	1	US-08-055-945-1	Sequence 1, Appli
	35	15.4	61.6	2508	4	US-08-981-392-1	Sequence 1, Appli
	36	15.4	61.6	2883	4	US-08-981-392-3	Sequence 3, Appli
c	37	15.2	60.8	50	2	US-08-850-049-122	Sequence 122, App
c	38	15.2	60.8	50	2	US-08-050-478-122	Sequence 122, App
c	39	15.2	60.8	50	4	US-09-414-117-122	Sequence 122, App
c	40	15.2	60.8	769	2	US-08-676-279-42	Sequence 42, Appl
c	41	15.2	60.8	1381	4	US-09-426-557-5	Sequence 5, Appli
c	42	15.2	60.8	1463	4	US-09-426-557-1	Sequence 1, Appli
c	43	15.2	60.8	1478	4	US-09-426-557-7	Sequence 7, Appli
c	44	15.2	60.8	1541	4	US-09-426-557-3	Sequence 3, Appli
c	45	15.2	60.8	1704	1	US-08-528-199-2	Sequence 2, Appli
c	46	15.2	60.8	1704	1	US-08-528-199-5	Sequence 5, Appli
c	47	15.2	60.8	31571	1	US-08-323-443B-1	Sequence 1, Appli
c	48	15.2	60.8	53526	3	US-08-658-136-2	Sequence 2, Appli
c	49	15.2	60.8	53577	3	US-08-658-136-1	Sequence 1, Appli
c	50	15	60.0	387	4	US-08-905-223-225	Sequence 225, App
c	51	15	60.0	666	4	US-08-998-416-995	Sequence 995, App
c	52	15	60.0	707	4	US-08-998-416-1093	Sequence 1093, Ap
c	53	15	60.0	1045	1	US-08-317-880-3	Sequence 3, Appli
c	54	15	60.0	1045	2	US-08-782-396-3	Sequence 3, Appli
c	55	15	60.0	1776	1	US-08-522-229B-1	Sequence 1, Appli
c	56	15	60.0	1776	2	US-09-102-433-1	Sequence 1, Appli
c	57	15	60.0	2426	4	US-08-528-026C-3	Sequence 3, Appli
c	58	15	60.0	3954	1	US-08-317-880-1	Sequence 1, Appli
c	59	15	60.0	3954	2	US-08-782-396-1	Sequence 1, Appli
c	60	14.8	59.2	605	4	US-09-385-982-515	Sequence 515, App
	61	14.8	59.2	1362	2	US-08-851-088-7	Sequence 7, Appli
	62	14.8	59.2	1362	3	US-08-851-089-1	Sequence 1, Appli
	63	14.8	59.2	1694	1	US-08-698-551-9	Sequence 9, Appli
	64	14.8	59.2	1694	2	US-08-602-228-9	Sequence 9, Appli
	65	14.8	59.2	1694	2	US-08-494-440B-9	Sequence 9, Appli
	66	14.8	59.2	1694	2	US-08-533-901B-9	Sequence 9, Appli
	67	14.8	59.2	1694	2	US-08-839-032A-9	Sequence 9, Appli
	68	14.8	59.2	1694	2	US-08-839-031A-9	Sequence 9, Appli
	69	14.8	59.2	1694	5	PCT-US95-12724-9	Sequence 9, Appli
c	70	14.8	59.2	2092	3	US-08-307-896-6	Sequence 6, Appli
c	71	14.8	59.2	2092	5	PCT-US95-11808-6	Sequence 6, Appli
	72	14.8	59.2	2473	1	US-08-698-551-17	Sequence 17, Appl
	73	14.8	59.2	2473	2	US-08-839-032A-17	Sequence 17, Appl
	74	14.8	59.2	4144	3	US-08-851-089-12	Sequence 12, Appl
c	75	14.8	59.2	4144	3	US-08-851-089-13	Sequence 13, Appl
	76	14.8	59.2	22108	3	US-09-053-197A-3	Sequence 3, Appli
	77	14.6	58.4	45	6	5474933-5	Patent No. 5474933
	78	14.6	58.4	632	4	US-09-328-111-798	Sequence 798, App
c	79	14.6	58.4	1182	1	US-08-241-943-23	Sequence 23, Appl
c	80	14.6	58.4	1431	1	US-08-254-357-1	Sequence 1, Appli
	81	14.6	58.4	1591	2	US-08-646-981-14	Sequence 14, Appl
c	82	14.6	58.4	2126	3	US-09-237-543-1	Sequence 1, Appli
	83	14.6	58.4	2322	4	US-08-624-635-4	Sequence 4, Appli
c	84	14.6	58.4	2327	6	5229279-5	Patent No. 5229279
c	85	14.6	58.4	2327	6	5512669-5	Patent No. 5512669
c	86	14.6	58.4	2463	1	US-08-339-578-1	Sequence 1, Appli
c	87	14.6	58.4	3243	1	US-08-611-107-32	Sequence 32, Appl
	88	14.6	58.4	3267	3	US-08-633-768A-3	Sequence 3, Appli
c	89	14.6	58.4	3868	3	US-09-357-070-1	Sequence 1, Appli
c	90	14.6	58.4	4154	1	US-08-131-365B-37	Sequence 37, Appl
c	91	14.6	58.4	4154	2	US-08-668-123-37	Sequence 37, Appl
c	92	14.6	58.4	4983	1	US-08-472-358-1	Sequence 1, Appli
c	93	14.6	58.4	4983	5	PCT-US92-05786A-1	Sequence 1, Appli
c	94	14.6	58.4	4984	1	US-08-687-806-1	Sequence 1, Appli
c	95	14.6	58.4	5220	2	US-08-777-405A-1	Sequence 1, Appli
c	96	14.6	58.4	5220	2	US-08-977-871A-1	Sequence 1, Appli
c	97	14.6	58.4	5220	2	US-09-225-951-1	Sequence 1, Appli
	98	14.6	58.4	5935	4	US-09-178-973B-17	Sequence 17, Appl
c	99	14.6	58.4	6763	2	US-08-756-506-23	Sequence 23, Appl
	100	14.6	58.4	6803	3	US-08-665-259-19	Sequence 19, Appl

ALIGNMENTS

```
RESULT   1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Enhanced Biotin Biosynthesis in Plant Tissue
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/351,970
;      FILING DATE:  08-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-401-068-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.0089;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db     105 CGATCCTCGTCAGGTGCAGGTCAGC 129


RESULT   2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-846-338-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.0089;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db     105 CGATCCTCGTCAGGTGCAGGTCAGC 129


RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A-15/9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1157
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 117
;      OTHER INFORMATION:  /product= "Biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioB"
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2295..3050
;      OTHER INFORMATION:  /codon_start= 2295
;      OTHER INFORMATION:  /function= "involved in pimeloyl-CoA synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /gene= "bioC"
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3750..5039
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3750
;      OTHER INFORMATION:  /EC_number= 2.6.1.62
;      OTHER INFORMATION:  /product= "DAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioA"
;      OTHER INFORMATION:  /number= 5
;      OTHER INFORMATION:  /standard_name=
;      OTHER INFORMATION:  "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
;      OTHER INFORMATION:  aminotransf."
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  5098..5574
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 5098
;      OTHER INFORMATION:  /function= "unknown, involved in biotin synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "ORFI"
;      OTHER INFORMATION:  /number= 6
;    FEATURE:
;      NAME/KEY:  -10_signal
;      LOCATION:  45..49
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  -35_signal
;      LOCATION:  23..28
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  105..119
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "bioB RBS no.9"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  2284..2297
;      OTHER INFORMATION:  /standard_name= "bioC RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3742..3752
;      OTHER INFORMATION:  /standard_name= "bioA RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  5088..5100
;      OTHER INFORMATION:  /standard_name= "ORFI RBS"
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  5583..5644
;      OTHER INFORMATION:  /standard_name= "rho-independent
;      OTHER INFORMATION:  transcriptional terminator"
;    FEATURE:
;      NAME/KEY:  stem_loop
;      LOCATION:  5583..5605
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..96
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "promoter ptac"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-1


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db     221 CGATCCTCGTCAGGTGCAGGTCAGC 245


RESULT   4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A15-9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1154..2308
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1154
;      OTHER INFORMATION:  /EC_number= 2.3.1.47
;      OTHER INFORMATION:  /product= "KAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "8-Amino-7-oxononanoate synthase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3043..3753
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3043
;      OTHER INFORMATION:  /EC_number= 6.3.3.3
;      OTHER INFORMATION:  /product= "DTB synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioD"
;      OTHER INFORMATION:  /number= 4
;      OTHER INFORMATION:  /standard_name= "Dethiobiotin synthase"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1141..1156
;      OTHER INFORMATION:  /standard_name= "bioF RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3030..3045
;      OTHER INFORMATION:  /standard_name= "bioD RBS"
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-6


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.011;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||||||||||||||||||||
Db     221 CGATCCTCGTCAGGTGCAGGTCAGC 245


RESULT   5
US-08-186-529-1
; Sequence 1, Application US/08186529
; Patent No. 5573764
;  GENERAL INFORMATION:
;    APPLICANT:  Sykes, Megan
;    APPLICANT:  Wolf, Stanley F.
;    TITLE OF INVENTION:  USE OF INTERLEUKIN-12 TO PREVENT
;    TITLE OF INVENTION:  GRAFT-VERSUS-HOST DISEASE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc., Legal Affairs
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/186,529
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McDaniels, Patricia A.
;      REGISTRATION NUMBER:  33,194
;      REFERENCE/DOCKET NUMBER:  GI 5225
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-498-8401
;      TELEFAX:  617-876-5851
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  987 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL TYPE:  Lymphoblast
;      CELL LINE:  RPMI 8866
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..987
US-08-186-529-1


  Query Match             72.8%;  Score 18.2;  DB 1;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 10;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
             || |||| |||||| ||||||||
Db     743 ATTCTCGGCAGGTGGAGGTCAGC 765


RESULT   6
US-08-640-386A-1
; Sequence 1, Application US/08640386A
; Patent No. 5756085
;  GENERAL INFORMATION:
;    APPLICANT:  Sykes, Megan
;    APPLICANT:  Wolf, Stanley F.
;    TITLE OF INVENTION:  USE OF INTERLEUKIN-12 TO PREVENT
;    TITLE OF INVENTION:  GRAFT-VERSUS-HOST DISEASE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc., Legal Affairs
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/640,386A
;       FILING DATE:
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Brown, Scott A.
;       REGISTRATION NUMBER:  32,724
;       REFERENCE/DOCKET NUMBER:  GI 5225A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-498-8224
;       TELEFAX:  617-876-5851
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  987 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;       CELL TYPE:  Lymphoblast
;       CELL LINE:  RPMI 8866
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..987
US-08-640-386A-1


  Query Match             72.8%;  Score 18.2;  DB 1;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 10;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 ATTCTCGGCAGGTGGAGGTCAGC 765


RESULT   7
US-08-848-760B-24
; Sequence 24, Application US/08848760B
; Patent No. 6248721
;   GENERAL INFORMATION:
;        APPLICANT: Chang, Lung-Ji
;        TITLE OF INVENTION: Animal Model For Evaluation Of Vaccines
;        NUMBER OF SEQUENCES: 33
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
;             STREET: 2421 N.W. 41st Street, Suite A-1
;             CITY: Gainesville
;             STATE: Florida
;             COUNTRY: United States of America
;             ZIP: 32606
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/848,760B
;             FILING DATE: 25-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/838,702
;             FILING DATE: 09-APR-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: PACE, DORAN R.
;             REGISTRATION NUMBER: 38,261
;             REFERENCE/DOCKET NUMBER: CNG-100C1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (352) 375-8100
;             TELEFAX: (352) 372-5800
;    INFORMATION FOR SEQ ID NO: 24:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 987 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: other nucleic acid
;             DESCRIPTION: /desc = "DNA"
;        SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-08-848-760B-24


  Query Match             72.8%;  Score 18.2;  DB 4;  Length 987;
  Best Local Similarity   87.0%;  Pred. No. 10;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 ATTCTCGGCAGGTGGAGGTCAGC 765


RESULT   8
US-08-184-009-194
; Sequence 194, Application US/08184009
; Patent No. 5833975
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;    NUMBER OF SEQUENCES:  217
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/184,009
;      FILING DATE:  19-JAN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2530
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066CURTMS
;  INFORMATION FOR SEQ ID NO:  194:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1018 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-184-009-194


  Query Match             72.8%;  Score 18.2;  DB 2;  Length 1018;
  Best Local Similarity   87.0%;  Pred. No. 10;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
```

```
Db    774 ATTCTCGGCAGGTGGAGGTCAGC 796


RESULT   9
US-08-458-356-194
; Sequence 194, Application US/08458356
; Patent No. 5942235
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;    NUMBER OF SEQUENCES:  217
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,356
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/184,009
;      FILING DATE:  19-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2530
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066CURTMS
;  INFORMATION FOR SEQ ID NO:  194:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1018 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-458-356-194


  Query Match             72.8%;  Score 18.2;  DB 2;  Length 1018;
  Best Local Similarity   87.0%;  Pred. No. 10;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     774 ATTCTCGGCAGGTGGAGGTCAGC 796


RESULT  10
US-08-460-736-194
; Sequence 194, Application US/08460736
; Patent No. 6265189
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;    NUMBER OF SEQUENCES:  217
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,736
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/184,009
;      FILING DATE:  19-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2530
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066CURTMS
;  INFORMATION FOR SEQ ID NO:  194:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1018 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-460-736-194


  Query Match             72.8%;  Score 18.2;  DB 4;  Length 1018;
  Best Local Similarity   87.0%;  Pred. No. 10;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     774 ATTCTCGGCAGGTGGAGGTCAGC 796


RESULT  11
US-08-751-767A-3
; Sequence 3, Application US/08751767A
; Patent No. 5994104
;  GENERAL INFORMATION:
;    APPLICANT:  ANDERSON, ROBERT J.
;    APPLICANT:  GRANT, HUGH
;    APPLICANT:  MACDONALD, IAN D.
;    TITLE OF INVENTION:  INTERLUKIN-12 FUSION PROTEIN
;    NUMBER OF SEQUENCES:  80
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/751,767A
;      FILING DATE:  08-NOV-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SADOFF, B.J.
;      REGISTRATION NUMBER:  36,663
;      REFERENCE/DOCKET NUMBER:  117-221
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  7038164091
;       TELEFAX:  7038164100
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1399 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  43..1026
US-08-751-767A-3


  Query Match            72.8%;  Score 18.2;  DB 2;  Length 1399;
  Best Local Similarity  87.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     785 ATTCTCGGCAGGTGGAGGTCAGC 807


RESULT  12
US-08-751-767A-11
; Sequence 11, Application US/08751767A
; Patent No. 5994104
;  GENERAL INFORMATION:
;     APPLICANT:  ANDERSON, ROBERT J.
;     APPLICANT:  GRANT, HUGH
;     APPLICANT:  MACDONALD, IAN D.
;     TITLE OF INVENTION:  INTERLUKIN-12 FUSION PROTEIN
;     NUMBER OF SEQUENCES:  80
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  NIXON & VANDERHYE P.C.
;       STREET:  1100 NORTH GLEBE ROAD
;       CITY:  ARLINGTON
;       STATE:  VA
;       COUNTRY:  USA
;       ZIP:  22201
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/751,767A
;       FILING DATE:  08-NOV-1996
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  SADOFF, B.J.
;       REGISTRATION NUMBER:  36,663
;       REFERENCE/DOCKET NUMBER:  117-221
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  7038164091
;       TELEFAX:  7038164100
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1560 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1557
US-08-751-767A-11


  Query Match            72.8%;  Score 18.2;  DB 2;  Length 1560;
  Best Local Similarity  87.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     680 ATTCTCGGCAGGTGGAGGTCAGC 702


RESULT  13
US-08-751-767A-9
; Sequence 9, Application US/08751767A
; Patent No. 5994104
;  GENERAL INFORMATION:
;     APPLICANT:  ANDERSON, ROBERT J.
;     APPLICANT:  GRANT, HUGH
;     APPLICANT:  MACDONALD, IAN D.
;     TITLE OF INVENTION:  INTERLUKIN-12 FUSION PROTEIN
;     NUMBER OF SEQUENCES:  80
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  NIXON & VANDERHYE P.C.
;       STREET:  1100 NORTH GLEBE ROAD
;       CITY:  ARLINGTON
;       STATE:  VA
;       COUNTRY:  USA
;       ZIP:  22201
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/751,767A
;       FILING DATE:  08-NOV-1996
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  SADOFF, B.J.
;       REGISTRATION NUMBER:  36,663
;       REFERENCE/DOCKET NUMBER:  117-221
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  7038164091
;       TELEFAX:  7038164100
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1623 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1620
US-08-751-767A-9


  Query Match            72.8%;  Score 18.2;  DB 2;  Length 1623;
  Best Local Similarity  87.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     743 ATTCTCGGCAGGTGGAGGTCAGC 765


RESULT  14
US-08-265-087-1
; Sequence 1, Application US/08265087
; Patent No. 5571515
;  GENERAL INFORMATION:
;     APPLICANT:  Scott, Phillip
;     APPLICANT:  Trinchieri, Giorgio
;     TITLE OF INVENTION:  Compositions and Methods for Use of
;     TITLE OF INVENTION:  IL-12 as an Adjuvant
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Howson and Howson
```

```
;       STREET:  Spring House Corporate Center, PO Box 457
;       CITY:  Spring House
;       STATE:  Pennsylvania
;       COUNTRY:  USA
;       ZIP:  19477
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/265,087
;       FILING DATE:
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/229,282
;       FILING DATE:  18-APR-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Bak, Mary E.
;       REGISTRATION NUMBER:  31,215
;       REFERENCE/DOCKET NUMBER:  WST51AUSA
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  215-540-9206
;       TELEFAX:  215-540-5818
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2362 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  33..1016
US-08-265-087-1


  Query Match             72.8%;  Score 18.2;  DB 1;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  15
US-08-621-493-1
; Sequence 1, Application US/08621493
; Patent No. 5723127
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, Phillip
;    APPLICANT:  Trinchieri, Giorgio
;    TITLE OF INVENTION:  Compositions and Methods for Use of
;    TITLE OF INVENTION:  IL-12 as an Adjuvant
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Center, PO Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/621,493
;      FILING DATE:  25-MAR-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/265,087
;      FILING DATE:  17-JUN-1994
;      APPLICATION NUMBER:  US 08/229,282
;      FILING DATE:  18-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  WST51AUSA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2362 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1016
US-08-621-493-1


  Query Match             72.8%;  Score 18.2;  DB 1;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| |||||| ||||||||
Db     775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  16
US-08-965-688-1
; Sequence 1, Application US/08965688
; Patent No. 5976539
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, Phillip
;    APPLICANT:  Trinchieri, Giorgio
;    TITLE OF INVENTION:  Compositions and Methods for Use of
;    TITLE OF INVENTION:  IL-12 as an Adjuvant
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Center, PO Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/965,688
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/621,493
;      FILING DATE:  25-MAR-1996
;      APPLICATION NUMBER:  08/265,087
;      FILING DATE:  17-JUN-1994
;      APPLICATION NUMBER:  US 08/229,282
;      FILING DATE:  18-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  WST51AUSA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
```

```
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2362 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1016
US-08-965-688-1


  Query Match             72.8%;  Score 18.2;  DB 2;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  17
US-09-260-173-1
; Sequence 1, Application US/09260173
; Patent No. 6168923
;  GENERAL INFORMATION:
;       APPLICANT: Scott, Phillip
;                  Trinchieri, Giorgio
;       TITLE OF INVENTION: Compositions and Methods for Use of
;                           IL-12 as an Adjuvant
;       NUMBER OF SEQUENCES: 4
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Howson and Howson
;            STREET: Spring House Corporate Center, PO Box 457
;            CITY: Spring House
;            STATE: Pennsylvania
;            COUNTRY: USA
;            ZIP: 19477
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/260,173
;            FILING DATE: 01-Mar-1999
;            CLASSIFICATION: <Unknown>
;                            17-JUN-1994
;                            18-APR-1994
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/965,688
;            FILING DATE: <Unknown>
;            APPLICATION NUMBER: 08/265,087
;            FILING DATE: 17-JUN-1994
;            APPLICATION NUMBER: US 08/229,282
;            FILING DATE: 18-APR-1994
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Bak, Mary E.
;            REGISTRATION NUMBER: 31,215
;            REFERENCE/DOCKET NUMBER: WST51AUSA
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 215-540-9206
;            TELEFAX: 215-540-5818
;  INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 2362 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: unknown
;       MOLECULE TYPE: cDNA
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  33..1016
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-260-173-1


  Query Match             72.8%;  Score 18.2;  DB 4;  Length 2362;
  Best Local Similarity   87.0%;  Pred. No. 11;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db    775 ATTCTCGGCAGGTGGAGGTCAGC 797


RESULT  18
US-08-751-767A-7
; Sequence 7, Application US/08751767A
; Patent No. 5994104
;  GENERAL INFORMATION:
;    APPLICANT:  ANDERSON, ROBERT J.
;    APPLICANT:  GRANT', HUGH
;    APPLICANT:  MACDONALD, IAN D.
;    TITLE OF INVENTION:  INTERLUKIN-12 FUSION PROTEIN
;    NUMBER OF SEQUENCES:  80
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/751,767A
;      FILING DATE:  08-NOV-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SADOFF, B.J.
;      REGISTRATION NUMBER:  36,663
;      REFERENCE/DOCKET NUMBER:  117-221
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  7038164091
;      TELEFAX:  7038164100
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6139 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3222..4841
US-08-751-767A-7


  Query Match             72.8%;  Score 18.2;  DB 2;  Length 6139;
  Best Local Similarity   87.0%;  Pred. No. 13;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcagc 25
          || |||| |||||| ||||||||
Db   3964 ATTCTCGGCAGGTGGAGGTCAGC 3986


RESULT  19
US-09-079-984A-1
; Sequence 1, Application US/09079984A
; Patent No. 6231850
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Okano, Fumiyoshi, Satoh, Massahiro,
;    APPLICANT:  Yamada, Katsushige
;    TITLE OF INVENTION:  Canine interleukin 12, a production method
;    TITLE OF INVENTION:  thereof, an immune disease treatment method and preventive
;    TITLE OF INVENTION:  method using it
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Miller & Christenbury Intellectual Property
;      ADDRESSEE:  Department of Schnader, Harrison, Segal and Lewis, LLP
;      STREET:  1600 Market Street, 39th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/079,984A
;      FILING DATE:  15-MAY-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Austin R. Miller
;      REGISTRATION NUMBER:  16,602
;      REFERENCE/DOCKET NUMBER:  1051-98
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-1810
;      TELEFAX:  (215) 568-6946
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  990 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;    FEATURE:
;      NAME/KEY:  Canine IL12
;      LOCATION:  1 to 987
;      IDENTIFICATION METHOD:  Similarity
US-09-079-984A-1


  Query Match             66.4%;  Score 16.6;  DB 4;  Length 990;
  Best Local Similarity   82.6%;  Pred. No. 53;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| || ||| ||||||||
Db     746 ATTCTCGGCACGTGGAGGTCAGC 768


RESULT  20
US-09-079-984A-11
; Sequence 11, Application US/09079984A
; Patent No. 6231850
;  GENERAL INFORMATION:
;    APPLICANT:  Okano, Fumiyoshi, Satoh, Massahiro,
;    APPLICANT:  Yamada, Katsushige
;    TITLE OF INVENTION:  Canine interleukin 12, a production method
;    TITLE OF INVENTION:  thereof, an immune disease treatment method and preventive
;    TITLE OF INVENTION:  method using it
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Miller & Christenbury Intellectual Property
;      ADDRESSEE:  Department of Schnader, Harrison, Segal and Lewis, LLP
;      STREET:  1600 Market Street, 39th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/079,984A
;      FILING DATE:  15-MAY-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Austin R. Miller
;      REGISTRATION NUMBER:  16,602
;      REFERENCE/DOCKET NUMBER:  1051-98
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 563-1810
;      TELEFAX:  (215) 568-6946
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  990 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;    FEATURE:
;      NAME/KEY:  Canine IL12
;      LOCATION:  1 to 987
;      IDENTIFICATION METHOD:  Similarity
US-09-079-984A-11


  Query Match             66.4%;  Score 16.6;  DB 4;  Length 990;
  Best Local Similarity   82.6%;  Pred. No. 53;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtcagc 25
           || |||| || ||| ||||||||
Db     746 ATTCTCGGCACGTGGAGGTCAGC 768


RESULT  21
US-09-049-672A-16/c
; Sequence 16, Application US/09049672A
; Patent No. 6135941
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Tang, Y. Tom
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Baughn, Mariah R.
;    TITLE OF INVENTION:  HUMAN IMMUNE SYSTEM ASSOCIATED PROTEINS
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/049,672A
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Cerrone, Michael C
;       REGISTRATION NUMBER:  39,132
;       REFERENCE/DOCKET NUMBER:  PF-0497 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-855-0555
;       TELEFAX:  650-845-4166
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  16:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3449 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;       LIBRARY:  BLADNOT04
;       CLONE:  1320068
US-09-049-672A-16


  Query Match             66.4%;  Score 16.6;  DB 3;  Length 3449;
  Best Local Similarity   82.6%;  Pred. No. 61;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcag 24
          | ||||||||||   ||||||||
Db    947 GGTCCTCGTCAGAGCCAGGTCAG 925


RESULT  22
US-08-175-471-6/c
; Sequence 6, Application US/08175471
; Patent No. 5665892
;  GENERAL INFORMATION:
;    APPLICANT:  Van Assche, C.
;    APPLICANT:  Lando, D.
;    APPLICANT:  Bruneau, J.M.
;    APPLICANT:  Voelker, T.
;    TITLE OF INVENTION:  Sucrase Phosphate Synthase (SPS), Its
;    TITLE OF INVENTION:  Process For Preparation, Its cDNA, and Utilization of cDNA To
;    TITLE OF INVENTION:  Modify The Expression Of SPS In The Plant Cells
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calgene, Inc.
;      STREET:  1920 Fifth Street
;      CITY:  Davis
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  95616
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.0 MB
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 6.0.7
;      SOFTWARE:  Microsoft Word 4.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/175,471
;      FILING DATE:  27-DEC-1993
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/672,646
;      FILING DATE:  18 MAR 1991
;      APPLICATION NUMBER:  EP 90402084.9
;      FILING DATE:  20 JUL 1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elizabeth Lassen
;      REGISTRATION NUMBER:  31,845
;      NAME:  Donna E. Scherer
;      REGISTRATION NUMBER:  34,719
;      REFERENCE/DOCKET NUMBER:  CGNE 72-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (916) 753-6313
;      TELEFAX:  (916) 753-1510
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3509 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
US-08-175-471-6


  Query Match             64.8%;  Score 16.2;  DB 1;  Length 3509;
  Best Local Similarity   85.7%;  Pred. No. 93;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtc 22
          ||||| |||| |||| |||||
Db    330 GATCCACGTCCGGTGGAGGTC 310


RESULT  23
US-08-429-054A-12/c
; Sequence 12, Application US/08429054A
; Patent No. 5917126
;  GENERAL INFORMATION:
;    APPLICANT:  VAN ASSCHE, CHARLES; LANDO, DANIELLE; BRUNEAU,
;    APPLICANT:  JEAN; VOELKER, TONI; GERVAIS, MONICA
;    TITLE OF INVENTION:  SUCROSE PHOSPHATE SYNTHETASE (SPS),
;    TITLE OF INVENTION:  A PREPARATION METHOD AND cDNA THEREFOR, AND USE OF THE
;    TITLE OF INVENTION:  cDNA FOR MODIFYING SPS EXPRESSION IN PLANT CELLS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIERMAN AND MUSERLIAN
;      STREET:  600 THIRD AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/429,054A
;      FILING DATE:  26-APR-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US  842,337
;      FILING DATE:  20-March-1992
;      APPLICATION NUMBER:  PCT/FR 91/00593
;      FILING DATE:  18-July-1991
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  French 90402094.9
;      FILING DATE:  20-July-1990
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles A. Muserlian
;      REGISTRATION NUMBER:  19,683
;      REFERENCE/DOCKET NUMBER:  146.1137
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 661-8000
;      TELEFAX:  (212) 661-8002
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3509
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Unknown
;      TOPOLOGY:  Unknown
;    MOLECULE TYPE:  cDNA
US-08-429-054A-12
```

```
  Query Match             64.8%;  Score 16.2;  DB 2;  Length 3509;
  Best Local Similarity   85.7%;  Pred. No. 93;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtc 22
          ||||| |||| |||| |||||
Db    330 GATCCACGTCCGGTGGAGGTC 310


RESULT  24
US-08-718-777-6/c
; Sequence 6, Application US/08718777
; Patent No. 5981852
;  GENERAL INFORMATION:
;    APPLICANT:  Van Assche, C.
;    APPLICANT:  Lando, D.
;    APPLICANT:  Bruneau, J. M.
;    APPLICANT:  Voelker, T.
;    APPLICANT:  Gervais, M.
;    TITLE OF INVENTION:  MODIFICATION OF SUCROSE
;    TITLE OF INVENTION:  PHOSPHATE
;    TITLE OF INVENTION:  SYNTHASE IN PLANTS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Barbara Rae-Venter
;      STREET:  260 Sheridan Avenue, Suite 440
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/718,777
;      FILING DATE:  NOT YET ASSIGNED
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/175,471
;      FILING DATE:  27-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barbara Rae-Venter
;      REGISTRATION NUMBER:  32,750
;      REFERENCE/DOCKET NUMBER:  CGNE.072.02US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)328-4400
;      TELEFAX:  (415)328-4477
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3509 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  112..3315
US-08-718-777-6


  Query Match             64.8%;  Score 16.2;  DB 2;  Length 3509;
  Best Local Similarity   85.7%;  Pred. No. 93;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtc 22
          ||||| |||| |||| |||||
Db    330 GATCCACGTCCGGTGGAGGTC 310


RESULT  25
US-09-078-862-2/c
; Sequence 2, Application US/09078862
; Patent No. 6091003
;  GENERAL INFORMATION:
;    APPLICANT:  Nan, Guo-Ling
;    APPLICANT:  Nagai, Chifumi
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR GENETIC
;    TITLE OF INVENTION:  TRANSFORMATION OF PINEAPPLE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/078,862
;      FILING DATE:  14-MAY-1998
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  UH-03321
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3509 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-078-862-2


  Query Match             64.8%;  Score 16.2;  DB 3;  Length 3509;
  Best Local Similarity   85.7%;  Pred. No. 93;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtc 22
          ||||| |||| |||| |||||
Db    330 GATCCACGTCCGGTGGAGGTC 310


RESULT  26
US-09-051-341-6/c
; Sequence 6, Application US/09051341
; Patent No. 6124528
;  GENERAL INFORMATION:
;    APPLICANT:  Shewmaker, C. K.
;    TITLE OF INVENTION:  MODIFICATION OF SOLUBLE SOLIDS USING
;    TITLE OF INVENTION:  SUCROSE PHOSPHATE SYNTHASE ENCODING SEQUENCE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rae-Venter Law Group, P.C.
;      STREET:  260 Sheridan Avenue, Suite 440
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/09/051,341
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/17351
;      FILING DATE:  25-OCT-1996
;      APPLICATION NUMBER:  US 08/549,016
;      FILING DATE:  27-OCT-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/372,200
;      FILING DATE:  12-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barbara Rae-Venter. Ph.D.,
;      REGISTRATION NUMBER:  32,750
;      REFERENCE/DOCKET NUMBER:  CGNE.110.02US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)328-4400
;      TELEFAX:  (415)328-4477
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3509 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  112..3315
US-09-051-341-6


  Query Match             64.8%;  Score 16.2;  DB 3;  Length 3509;
  Best Local Similarity   85.7%;  Pred. No. 93;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtc 22
           ||||| |||| |||| |||||
Db     330 GATCCACGTCCGGTGGAGGTC 310


RESULT  27
US-08-336-408B-3/c
; Sequence 3, Application US/08336408B
; Patent No. 5723329
;  GENERAL INFORMATION:
;    APPLICANT:  EVANS, RONALD M.
;    APPLICANT:  MANGELSDORF, DAVID J.
;    TITLE OF INVENTION:  RETINOID RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
;      STREET:  444 SOUTH FLOWER STREET, SUITE 2000
;      CITY:  LOS ANGELES
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/336,408B
;      FILING DATE:  08-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/933,453
;      FILING DATE:  21-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US91/00399
;      FILING DATE:  22-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/478,071
;      FILING DATE:  09-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  REITER, STEPHEN E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9851
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-1995
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2095 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  178..1581
US-08-336-408B-3


  Query Match             63.2%;  Score 15.8;  DB 1;  Length 2095;
  Best Local Similarity   89.5%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 ctcgtcaggtgcaggtcag 24
           |||| |||||| |||||||
Db     709 CTCGGCAGGTGTAGGTCAG 691


RESULT  28
PCT-US91-00399-3/c
; Sequence 3, Application PC/TUS9100399
;  GENERAL INFORMATION:
;    APPLICANT:  Mangelsdorf, Dr., David J.
;    APPLICANT:  Evans Dr., Ronald M.
;    TITLE OF INVENTION:  RETINOID RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FITCH, EVEN, TABIN & FLANNERY
;      STREET:  135 South LaSalle Street, Suite 900
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/00399
;      FILING DATE:  19910122
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/478,071
;      FILING DATE:  09-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Mr., Stephen E.
;      REGISTRATION NUMBER:  31192
;      REFERENCE/DOCKET NUMBER:  50852
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 552-1311
;      TELEFAX:  (619) 552-0095
;      TELEX:  20 6566 PATLAW CGO
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2095 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;     MOLECULE TYPE:  DNA (genomic)
;     IMMEDIATE SOURCE:
;       CLONE:  RXR MOUSE ALPHA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  join(178..1581)
PCT-US91-00399-3


  Query Match             63.2%;  Score 15.8;  DB 5;  Length 2095;
  Best Local Similarity   89.5%;  Pred. No. 1.3e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 ctcgtcaggtgcaggtcag 24
           |||| |||||| |||||||
Db     709 CTCGGCAGGTGTAGGTCAG 691


RESULT  29
5457037-1/c
;Patent No. 5457037
;     APPLICANT: TOGNONI, ANGELO;CARRERA, PAOLO;CAMERINI, BARBARA;
;GALLI, GIULIANO;LUCCHESE, GIUSEPPE;GRANDI, GUIDO;DI GENNARO, CARLO
;     TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOAMYLASE
; ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME
;     NUMBER OF SEQUENCES: 7
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/1,797
;       FILING DATE: 08-JAN-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 749,621
;       FILING DATE: 19-AUG-1991
;       APPLICATION NUMBER: 224,114
;       FILING DATE: 25-JUL-1988
;SEQ ID NO:1:
;       LENGTH: 3336
5457037-1


  Query Match             62.4%;  Score 15.6;  DB 6;  Length 3336;
  Best Local Similarity   81.8%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtc 22
           ||||| ||  |||||||| |||
Db      81 CGATCATCCACAGGTGCACGTC 60


RESULT  30
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 80246
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4


  Query Match             62.4%;  Score 15.6;  DB 4;  Length 80246;
  Best Local Similarity   81.8%;  Pred. No. 2.4e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtca 23
           ||||| || || |||||| |||
Db   31843 gatccccggcatgtgcagttca 31864


RESULT  31
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 80595
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3


  Query Match             62.4%;  Score 15.6;  DB 4;  Length 80595;
  Best Local Similarity   81.8%;  Pred. No. 2.4e+02;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtca 23
           ||||| || || |||||| |||
Db   32098 gatccccggcatgtgcagttca 32119


RESULT  32
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


  Query Match             62.4%;  Score 15.6;  DB 4;  Length 4403765;
  Best Local Similarity   87.0%;  Pred. No. 49;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           ||||| |||||||||||| | |
Db 3643730 gatccgcgtcaggtgcaggactg 3643752


RESULT  33
US-09-103-840A-1
```

```
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1


  Query Match             62.4%;  Score 15.6;  DB 4;  Length 4411529;
  Best Local Similarity   87.0%;  Pred. No. 49;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           ||||| |||||||||||||| | |
Db 3648880 gatccgcgtcaggtgcaggactg 3648902


RESULT  34
US-08-055-945-1/c
; Sequence 1, Application US/08055945
; Patent No. 5462855
;  GENERAL INFORMATION:
;    APPLICANT:  Springer, Wolfgang; Rast, Hans George;
;    APPLICANT:  L bberding, Antonius; and Kanne, Reinhard
;    TITLE OF INVENTION:  A METHOD FOR THE DETECTION AND
;    TITLE OF INVENTION:  QUANTITATIVE DETERMINATION OF
;    TITLE OF INVENTION:  NITROSOMONAS STRAINS IN
;    TITLE OF INVENTION:  WASTEWATERS OR SOILS
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SPRUNG HORN KRAMER & WOODS
;      STREET:  1140 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 1.2 MB
;      MEDIUM TYPE:  storage
;      COMPUTER:  NEC Powermate 1 Plus
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/055,945
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/692,745
;      FILING DATE:  No. 5462855e
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  Bayer 8194-KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 391-0520
;      TELEFAX:  (212) 382-0949
;      TELEX:  423092 NYP UI
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1722 bp
;      TYPE:  Nucleotide
;      STRANDEDNESS:  Double
;      TOPOLOGY:  Linear
;    MOLECULE TYPE:  Genomic DNA
US-08-055-945-1


  Query Match             61.6%;  Score 15.4;  DB 1;  Length 1722;
  Best Local Similarity   76.0%;  Pred. No. 2e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           |||||||   |||||| | |||| |
Db     970 CGATCCTGAACAGGTGGATGTCAAC 946


RESULT  35
US-08-981-392-1
; Sequence 1, Application US/08981392
; Patent No. 6262025
;  GENERAL INFORMATION:
;    APPLICANT:  Ish-Horowicz, David
;    APPLICANT:  Henrique, Domingos Manuel Pinto
;    APPLICANT:  Lewis, Julian Hart
;    APPLICANT:  Artavanis-Tsakonas, Spyridon
;    APPLICANT:  Gray, Grace
;    TITLE OF INVENTION:  NUCLEOTIDE AND PROTEIN SEQUENCES
;    TITLE OF INVENTION:  OF VERTEBRATE DELTA GENES AND METHODS BASED THEREON
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036/2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/981,392
;      FILING DATE:  22-DEC-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Antler, Adriane M.
;      REGISTRATION NUMBER:  32,605
;      REFERENCE/DOCKET NUMBER:  7326-038
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2508 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  277...2460
;      OTHER INFORMATION:
US-08-981-392-1


  Query Match             61.6%;  Score 15.4;  DB 4;  Length 2508;
  Best Local Similarity   76.0%;  Pred. No. 2e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
```

```
          ||| || | ||||| || ||||| |
Db   1860 CGAGCCACCTCAGGGGCCGGTCATC 1884


RESULT  36
US-08-981-392-3
; Sequence 3, Application US/08981392
; Patent No. 6262025
;  GENERAL INFORMATION:
;    APPLICANT:  Ish-Horowicz, David
;    APPLICANT:  Henrique, Domingos Manuel Pinto
;    APPLICANT:  Lewis, Julian Hart
;    APPLICANT:  Artavanis-Tsakonas, Spyridon
;    APPLICANT:  Gray, Grace
;    TITLE OF INVENTION:  NUCLEOTIDE AND PROTEIN SEQUENCES
;    TITLE OF INVENTION:  OF VERTEBRATE DELTA GENES AND METHODS BASED THEREON
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036/2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/981,392
;      FILING DATE:  22-DEC-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Antler, Adriane M.
;      REGISTRATION NUMBER:  32,605
;      REFERENCE/DOCKET NUMBER:  7326-038
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2883 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-981-392-3


  Query Match             61.6%;  Score 15.4;  DB 4;  Length 2883;
  Best Local Similarity   76.0%;  Pred. No. 2.1e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
           ||| || | ||||| || ||||| |
Db    1860 CGAGCCACCTCAGGGGCCGGTCATC 1884


RESULT  37
US-08-850-049-122/c
; Sequence 122, Application US/08850049
; Patent No. 5965726
;  GENERAL INFORMATION:
;    APPLICANT:
;    APPLICANT:
;    APPLICANT:
;    TITLE OF INVENTION:  METHOD OF ELIMINATING
;    TITLE OF INVENTION:  INHIBITORY/INSTABILITY REGIONS OF mRNA
;    NUMBER OF SEQUENCES:  130
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/850,049
;      FILING DATE:  02-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/050,478
;      FILING DATE:  26-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/02908
;      FILING DATE:  29-MAR-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/858,747
;      FILING DATE:  27-MAR-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MORRY, MARY J.
;      REGISTRATION NUMBER:  34,398
;      REFERENCE/DOCKET NUMBER:  2026-4006US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)758-4800
;      TELEFAX:  (212)751-6849
;  INFORMATION FOR SEQ ID NO:  122:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  50 BASE PAIRS
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
US-08-850-049-122


  Query Match             60.8%;  Score 15.2;  DB 2;  Length 50;
  Best Local Similarity   85.0%;  Pred. No. 1.6e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcagg 20
           |||||||||||| |  ||||
Db      43 CGATCCTCGTCACGATCAGG 24


RESULT  38
US-08-050-478-122/c
; Sequence 122, Application US/08050478
; Patent No. 5972596
;  GENERAL INFORMATION:
;    APPLICANT:
;    APPLICANT:
;    APPLICANT:
;    TITLE OF INVENTION:  METHOD OF ELIMINATING
;    TITLE OF INVENTION:  INHIBITORY/INSTABILITY REGIONS OF mRNA
;    NUMBER OF SEQUENCES:  130
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/050,478
;      FILING DATE:  26-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/02908
;      FILING DATE:  29-MAR-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/858,747
;      FILING DATE:  27-MAR-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MORRY, MARY J.
;      REGISTRATION NUMBER:  34,398
;      REFERENCE/DOCKET NUMBER:  2026-4006US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)758-4800
;      TELEFAX:  (212)751-6849
;  INFORMATION FOR SEQ ID NO:  122:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  50 BASE PAIRS
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
US-08-050-478-122


  Query Match             60.8%;  Score 15.2;  DB 2;  Length 50;
  Best Local Similarity   85.0%;  Pred. No. 1.6e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcagg 20
          ||||||||||||  |  ||||
Db     43 CGATCCTCGTCACGATCAGG 24


RESULT  39
US-09-414-117-122/c
; Sequence 122, Application US/09414117
; Patent No. 6291664
;  GENERAL INFORMATION:
;    APPLICANT:
;    APPLICANT:
;    APPLICANT:
;    TITLE OF INVENTION:  METHOD OF ELIMINATING
;    TITLE OF INVENTION:  INHIBITORY/INSTABILITY REGIONS OF mRNA
;    NUMBER OF SEQUENCES:  130
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/414,117
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/850,049
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/02908
;      FILING DATE:  29-MAR-1993
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/858,747
;      FILING DATE:  27-MAR-1992
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MORRY, MARY J.
;      REGISTRATION NUMBER:  34,398
;      REFERENCE/DOCKET NUMBER:  2026-4006US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)758-4800
;      TELEFAX:  (212)751-6849
;  INFORMATION FOR SEQ ID NO:  122:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  50 BASE PAIRS
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
US-09-414-117-122


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 50;
  Best Local Similarity   85.0%;  Pred. No. 1.6e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcagg 20
          ||||||||||||  |  ||||
Db     43 CGATCCTCGTCACGATCAGG 24


RESULT  40
US-08-676-279-42/c
; Sequence 42, Application US/08676279
; Patent No. 5869247
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  MACROPHAGE NUCLEOTIDE SEQUENCE
;    NUMBER OF SEQUENCES:  63
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,279
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB95/00095
;      APPLICATION NUMBER:  GB 9400929.7
;      FILING DATE:  19-JAN-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9422021.7
;      FILING DATE:  31-OCT-1994
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  769 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-676-279-42


  Query Match             60.8%;  Score 15.2;  DB 2;  Length 769;
  Best Local Similarity   85.0%;  Pred. No. 2.2e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      6 ctcgtcaggtgcaggtcagc 25
          || | ||||| |||||||||
Db    253 CTGGGCAGGTTCAGGTCAGC 234


RESULT  41
US-09-426-557-5/c
; Sequence 5, Application US/09426557
```

```
; Patent No. 6232527
; GENERAL INFORMATION:
;  APPLICANT: Mahajan, Pramod B.
;  TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses
;  TITLE OF INVENTION: Thereof
;  FILE REFERENCE: 0961
;  CURRENT APPLICATION NUMBER: US/09/426,557
;  CURRENT FILING DATE: 1999-10-22
;  EARLIER APPLICATION NUMBER: 60/112,332
;  EARLIER FILING DATE: 1998-12-15
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 1381
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (37)...(1173)
US-09-426-557-5


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 1381;
  Best Local Similarity   85.0%;  Pred. No. 2.3e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 tcctcgtcaggtgcaggtca 23
           ||||| |||||||||||  ||
Db     947 TCCTCATCAGGTGCAGTCCA 928


RESULT  42
US-09-426-557-1/c
; Sequence 1, Application US/09426557
; Patent No. 6232527
; GENERAL INFORMATION:
;  APPLICANT: Mahajan, Pramod B.
;  TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses
;  TITLE OF INVENTION: Thereof
;  FILE REFERENCE: 0961
;  CURRENT APPLICATION NUMBER: US/09/426,557
;  CURRENT FILING DATE: 1999-10-22
;  EARLIER APPLICATION NUMBER: 60/112,332
;  EARLIER FILING DATE: 1998-12-15
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 1463
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (85)...(1221)
US-09-426-557-1


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 1463;
  Best Local Similarity   85.0%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 tcctcgtcaggtgcaggtca 23
           ||||| |||||||||||  ||
Db     995 TCCTCATCAGGTGCAGTCCA 976


RESULT  43
US-09-426-557-7/c
; Sequence 7, Application US/09426557
; Patent No. 6232527
; GENERAL INFORMATION:
;  APPLICANT: Mahajan, Pramod B.
;  TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses
;  TITLE OF INVENTION: Thereof
;  FILE REFERENCE: 0961
;  CURRENT APPLICATION NUMBER: US/09/426,557
;  CURRENT FILING DATE: 1999-10-22
;  EARLIER APPLICATION NUMBER: 60/112,332
;  EARLIER FILING DATE: 1998-12-15
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
;   LENGTH: 1478
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (97)...(1233)
US-09-426-557-7


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 1478;
  Best Local Similarity   85.0%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 tcctcgtcaggtgcaggtca 23
           ||||| |||||||||||  ||
Db    1007 TCCTCATCAGGTGCAGTCCA 988


RESULT  44
US-09-426-557-3/c
; Sequence 3, Application US/09426557
; Patent No. 6232527
; GENERAL INFORMATION:
;  APPLICANT: Mahajan, Pramod B.
;  TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses
;  TITLE OF INVENTION: Thereof
;  FILE REFERENCE: 0961
;  CURRENT APPLICATION NUMBER: US/09/426,557
;  CURRENT FILING DATE: 1999-10-22
;  EARLIER APPLICATION NUMBER: 60/112,332
;  EARLIER FILING DATE: 1998-12-15
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 1541
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (79)...(1215)
US-09-426-557-3


  Query Match             60.8%;  Score 15.2;  DB 4;  Length 1541;
  Best Local Similarity   85.0%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 tcctcgtcaggtgcaggtca 23
           ||||| |||||||||||  ||
Db     989 TCCTCATCAGGTGCAGTCCA 970


RESULT  45
US-08-528-199-2/c
; Sequence 2, Application US/08528199
; Patent No. 5763228
;  GENERAL INFORMATION:
;    APPLICANT:  KUBOTA, Michio
;    APPLICANT:  TSUSAKI, Keiji
;    APPLICANT:  SUGIMOTO, Toshiyuki
;    TITLE OF INVENTION:  RECOMBINANT ENZYME FOR CONVERTING
;    TITLE OF INVENTION:  MALTOSE INTO TREHALOSE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
```

```
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/528,199
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/485,126
;      FILING DATE:  14-SEP-1995
;      APPLICATION NUMBER:  US 08/485,126
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 156399/1994
;      FILING DATE:  16-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROWDY, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  KUBOTA=6A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1704 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-528-199-2


  Query Match             60.8%;  Score 15.2;  DB 1;  Length 1704;
  Best Local Similarity   85.0%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggt 21
            |||| ||| |||||||||| |
Db    1293 GATCGTCGGCAGGTGCAGCT 1274



Search completed: December 26, 2001, 12:28:38
Job time: 7475 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:12:09 ; Search time 4612.2 Seconds
(without alignments)
58.246 Million cell updates/sec

Title: US-09-396-196F-9
Perfect score: 25
Sequence: 1 cgatcctcgtcaggtgcaggtcagc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	18.8	75.2	543	13	AZ168021	AZ168021 SP_0103_B
c	2	18.2	72.8	300	10	AU100606	AU100606 AU100606
c	3	18.2	72.8	354	10	AW826993	AW826993 fk55g12.y
c	4	18.2	72.8	371	10	AW143033	AW143033 EST293328
c	5	18.2	72.8	461	13	AQ846280	AQ846280 LMAJFV1_1
	6	18.2	72.8	872	13	CNS026TZ	AL183824 Tetraodon
c	7	18.2	72.8	883	10	AL556337	AL556337 AL556337
c	8	18.2	72.8	904	11	BG023733	BG023733 602303347
c	9	18.2	72.8	917	10	AL580709	AL580709 AL580709
c	10	18.2	72.8	1095	11	BF979422	BF979422 602288251
c	11	17.8	71.2	494	13	AQ982558	AQ982558 RPCI-23-2
	12	17.8	71.2	536	10	BE234470	BE234470 141586 MA
	13	17.8	71.2	777	10	AL538687	AL538687 AL538687
	14	17.8	71.2	1032	13	CNS044UQ	AL274571 Tetraodon
c	15	17.8	71.2	1064	11	BI078629	BI078629 602872730
	16	17.6	70.4	339	10	AA594461	AA594461 nl94c07.s
c	17	17.6	70.4	360	11	BF318689	BF318689 uy05a01.y
	18	17.6	70.4	369	11	BF659032	BF659032 maa89h11.
	19	17.6	70.4	398	11	BE860402	BE860402 UI-M-AN1-
c	20	17.6	70.4	413	10	AA644735	AA644735 M-EST008
c	21	17.6	70.4	413	11	BE862745	BE862745 UI-M-BH0-
c	22	17.6	70.4	424	13	AQ172663	AQ172663 HS_3194_A
c	23	17.6	70.4	429	13	AQ371411	AQ371411 HS_5043_A
	24	17.6	70.4	463	11	BE850719	BE850719 ux01a08.y
c	25	17.6	70.4	495	10	AI789420	AI789420 uk99a06.y
	26	17.6	70.4	508	10	AI180705	AI180705 ub91b01.r
	27	17.6	70.4	574	10	AU204314	AU204314 AU204314
	28	17.6	70.4	618	10	AA536738	AA536738 vj88c01.r
	29	17.6	70.4	628	10	BE653574	BE653574 UI-M-AM1-
c	30	17.6	70.4	641	13	AZ426129	AZ426129 1M0206H03
c	31	17.6	70.4	721	10	BE262905	BE262905 601153594
c	32	17.6	70.4	833	11	BG478947	BG478947 602525709
c	33	17.6	70.4	893	11	BF794272	BF794272 602255473
	34	17.6	70.4	897	13	CNS02FRJ	AL195400 Tetraodon
c	35	17.6	70.4	929	11	BF313136	BF313136 601897654
	36	17.6	70.4	942	13	CNS00L1P	AL067857 Drosophil
c	37	17.6	70.4	982	13	CNS006PE	AL065656 Drosophil
c	38	17.6	70.4	985	13	CNS00HWP	AL073836 Drosophil
c	39	17.6	70.4	1002	11	BF965985	BF965985 602277168
c	40	17.6	70.4	1016	13	CNS00HQI	AL074118 Drosophil
c	41	17.4	69.6	1126	11	BE795069	BE795069 601592885
c	42	17.4	69.6	1298	11	BG328084	BG328084 602427134
	43	17.2	68.8	108	10	AW371554	AW371554 QV3-BT029
c	44	17.2	68.8	112	10	AA651689	AA651689 nn47b06.r
c	45	17.2	68.8	235	10	AA197041	AA197041 zq10c11.r
	46	17.2	68.8	236	11	BF349456	BF349456 RC2-HT014
c	47	17.2	68.8	255	11	N83353	N83353 EC0312F Hum
c	48	17.2	68.8	261	11	T98142	T98142 ye30g07.r1
c	49	17.2	68.8	284	11	BE830330	BE830330 RC5-ET008
c	50	17.2	68.8	298	10	AI766284	AI766284 wh71b04.x
c	51	17.2	68.8	330	11	T94973	T94973 ye38f05.r1
	52	17.2	68.8	331	11	BI042497	BI042497 RC5-OT009
	53	17.2	68.8	340	10	AA366219	AA366219 EST77351
c	54	17.2	68.8	362	11	T94710	T94710 ye35d10.r1
c	55	17.2	68.8	378	10	AW905010	AW905010 RC5-NN106
	56	17.2	68.8	379	11	BE955423	BE955423 UI-M-BH4-
c	57	17.2	68.8	380	10	AA112811	AA112811 zn52c03.r
c	58	17.2	68.8	380	11	N80501	N80501 za10a05.r1
	59	17.2	68.8	385	11	BG990391	BG990391 MR2-HT116
c	60	17.2	68.8	387	10	AA094761	AA094761 cp1441.se
c	61	17.2	68.8	395	11	BF815067	BF815067 MR2-CI012
c	62	17.2	68.8	413	11	N80543	N80543 za10g05.r1
c	63	17.2	68.8	417	11	W24620	W24620 zb62c08.r1
c	64	17.2	68.8	425	10	AA569432	AA569432 nf22e03.s
c	65	17.2	68.8	431	10	AA147333	AA147333 zo38h07.r
c	66	17.2	68.8	438	10	AA188286	AA188286 zp71b11.r
c	67	17.2	68.8	438	11	T59384	T59384 yb57e10.r1
c	68	17.2	68.8	439	11	H95000	H95000 yw57c04.r1
c	69	17.2	68.8	445	10	AA493685	AA493685 nh01g01.s
c	70	17.2	68.8	453	10	AA485986	AA485986 ab38b04.r
c	71	17.2	68.8	471	11	W44474	W44474 zc29d11.r1
	72	17.2	68.8	478	13	AZ345822	AZ345822 1M0080M08
c	73	17.2	68.8	479	11	BG820193	BG820193 602782240
c	74	17.2	68.8	479	11	W46865	W46865 zc38d08.r1
c	75	17.2	68.8	481	11	W32788	W32788 zc69a03.r1
	76	17.2	68.8	485	10	AA040207	AA040207 zk42b12.r
c	77	17.2	68.8	486	10	AA987346	AA987346 or92e10.s
c	78	17.2	68.8	492	10	AA580579	AA580579 nd42c04.r
c	79	17.2	68.8	496	10	AW161475	AW161475 au71c11.y
c	80	17.2	68.8	501	10	AA634987	AA634987 ab47b03.r
c	81	17.2	68.8	519	11	BF378097	BF378097 PM3-TN015
c	82	17.2	68.8	520	10	AA432048	AA432048 zw80f08.r
c	83	17.2	68.8	528	11	N24253	N24253 yx22g11.r1
c	84	17.2	68.8	529	11	T70108	T70108 yc17b05.r1
c	85	17.2	68.8	535	10	AA532884	AA532884 nj59d08.s

c 86	17.2	68.8	544	13	AZ143198	AZ143198 SP_0028_B
87	17.2	68.8	549	10	BE013697	BE013697 125147 MA
c 88	17.2	68.8	555	11	BF266090	BF266090 HV_CEa001
c 89	17.2	68.8	561	10	AA081071	AA081071 zn32f12.r
c 90	17.2	68.8	580	10	AA626485	AA626485 ab50a10.r
c 91	17.2	68.8	582	10	AA829288	AA829288 of08a03.s
c 92	17.2	68.8	593	10	AA580493	AA580493 nn22b02.s
c 93	17.2	68.8	607	10	AA533254	AA533254 nj66b05.s
94	17.2	68.8	617	10	BE739491	BE739491 601556410
c 95	17.2	68.8	622	13	AZ863344	AZ863344 2M0171L17
c 96	17.2	68.8	627	10	BE392606	BE392606 601307432
c 97	17.2	68.8	631	10	BE378226	BE378226 601238009
c 98	17.2	68.8	639	11	BG766460	BG766460 602739220
99	17.2	68.8	641	10	BE619666	BE619666 601472807
c 100	17.2	68.8	645	10	BE388542	BE388542 601281825

ALIGNMENTS

```
RESULT   1
AZ168021/c
LOCUS       AZ168021      543 bp    DNA             GSS       29-AUG-2000
DEFINITION  SP_0103_B1_F05_T7A Strongylocentrotus purpuratus, purple sea urchin
            , sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=103 Col=9 Row=L, DNA sequence.
ACCESSION   AZ168021
VERSION     AZ168021.1  GI:8338389
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 103  row: L  column: 9
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 543.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="Plate=103 Col=9 Row=L"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
BASE COUNT      149 a    116 c    116 g    157 t      5 others
ORIGIN


  Query Match             75.2%;  Score 18.8;  DB 13;  Length 543;
  Best Local Similarity   90.9%;  Pred. No. 6.7e+02;
  Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 atcctcgtcaggtgcaggtcag 24
          |||||||||||||||| ||| ||
```

```
Db    115 ATCCTCGTCAGGTGCTGGTAAG 94


RESULT   2
AU100606/c
LOCUS       AU100606      300 bp    mRNA            EST       05-APR-2001
DEFINITION  AU100606 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            y79a3N51 similar to Homo sapiens mRNA for cystinosin, mRNA
            sequence.
ACCESSION   AU100606
VERSION     AU100606.1  GI:13551735
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
            ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
            ,K., Suyama,A. and Sugano,S.
  TITLE     In silico mapping of the 5'-ends of human mRNAs using full-length
            enriched and 5'-end enriched cDNA libraries constructed by
            Oligo-capping method
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="y79a3N51"
                     /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT       51 a     93 c     97 g     51 t      8 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 300;
  Best Local Similarity   87.0%;  Pred. No. 1.1e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcag 24
          || ||||| ||||||| ||||||
Db     62 GAGCCTCGCCAGGTGCCGGTCAG 40


RESULT   3
AW826993/c
LOCUS       AW826993      354 bp    mRNA            EST       17-MAY-2000
DEFINITION  fk55g12.y1 Zebrafish adult retina cDNA Danio rerio cDNA 5' similar
            to TR:Q15846 Q15846 HYPOTHETICAL 54.2 KD PROTEIN PRECURSOR. ;, mRNA
            sequence.
ACCESSION   AW826993
VERSION     AW826993.1  GI:7920070
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 354)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
```

```
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by: Susan E. Brockerhoff DNA Sequencing by:
            Washington University Genome Sequencing Center Clone distribution:
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 292.
FEATURES             Location/Qualifiers
     source          1. .354
                     /organism="Danio rerio"
                     /strain="wild-type"
                     /db_xref="taxon:7955"
                     /clone_lib="Zebrafish adult retina cDNA"
                     /sex="mixed"
                     /dev_stage="1-2 years"
                     /lab_host="E.Coli XL1-Blue MRF' (XL1-Blue MRF')"
                     /note="Vector: Lambda ZAP II (pBluescript SK-); Site_1:
                     EcoRI; Site_2: SalI; This Zebrafish library was
                     constructed by Dr. Susan E. Brockerhoff (email:
                     sbrocker@u.washington.edu) RZPD library number: 760"
BASE COUNT       83 a    102 c     81 g     87 t      1 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 354;
  Best Local Similarity   83.3%;  Pred. No. 1.1e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcag 24
           |||||||| ||  ||||||||||
Db     235 CGATCCTCACCACNTGCAGGTCAG 212


RESULT   4
AW143033/c
LOCUS       AW143033      371 bp    mRNA            EST       30-OCT-1999
DEFINITION  EST293328 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
            RGIBC08 5' end, mRNA sequence.
ACCESSION   AW143033
VERSION     AW143033.1  GI:6162932
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 371)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST please check the TIGR Rat Gene
            Index (http://www.tigr.org/tdb/rgi/rgi.html). To order a clone
            contact the ATCC (http://www.atcc.org/atcc.html).
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .371
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
                     /clone="RGIBC08"
                     /clone_lib="Normalized rat kidney, Bento Soares"
                     /note="Organ: kidney; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
BASE COUNT       82 a    117 c    103 g     69 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 371;
  Best Local Similarity   87.0%;  Pred. No. 1.1e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           | ||||| |||||||||||| |||
Db     212 GCTCCTCCTCAGGTGCAGGGCAG 190


RESULT   5
AQ846280/c
LOCUS       AQ846280      461 bp    DNA             GSS       25-MAY-2001
DEFINITION  LMAJFV1_lm17b03.y1 Leishmania major FV1 random genomic library
            Leishmania major genomic clone LMAJFV1_lm17b03 5', DNA sequence.
ACCESSION   AQ846280
VERSION     AQ846280.1  GI:6050928
KEYWORDS    GSS.
SOURCE      Leishmania major.
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Akopyants,N.S., Clifton,S.W., Martin,J., Pape,D., Wylie,T., Li,L.,
            Kissinger,J., Roos,D.S., Marra,M., Hillier,L., Chinwalla,A.,
            Blistain,A., Schmitt,A., Person,B., Theising,B., Ritter,E., Ronko
            ,I., Bennett,J., Cole,R., Underwood,K., Cardenas,M., Gibbons,M.,
            Harvey,N., McCann,R., Tsagareishvili,R., Williams,T., Jackson,Y.,
            Bowers,Y., Swaller,T., Waterston,R., Wilson,R. and Beverley,S.M.
  TITLE     A survey of the Leishmania major Friedlin strain V1 genome by
            shotgun sequencing: a resource for DNA microarrays and expression
            profiling
  JOURNAL   Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
  MEDLINE   21192569
COMMENT     Other_GSSs: lm17b03.x1
            Contact: Akopyants, NS / Beverley, SM
            WashU Leishmania Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library construction: Natalia S. Akopyants, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             If using this information please cite:
             N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
            Friedlin strain V1 genome by shotgun sequencing' and the Washington
            University Genome Sequencing Center For information on obtaining
            clone material please contact: Natalia S. Akopyants Ph.D.
            (natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
            (beverley@borcim.wustl.edu)
            Seq primer: -40RP from Gibco
            Class: shotgun
            High quality sequence stop: 373.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Leishmania major"
                     /strain="Friedlin strain V1"
                     /db_xref="taxon:5664"
                     /clone="LMAJFV1_lm17b03"
                     /clone_lib="Leishmania major FV1 random genomic library"
                     /lab_host="TOP10 (Invitrogen)"
                     /note="Vector: pZero-2 (Invitrogen); Site_1: EcoRV;
                     Genomic DNA was isolated from stationary phase cells. For
                     this library, DNA was sheared to give a tight size
                     distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
```

polymerase, dephosphorylated with Shrimp Alkaline
Phosphatase and ligated into pZero-2 vector's EcoRV site."
BASE COUNT 103 a 131 c 139 g 88 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 461;
Best Local Similarity 87.0%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gatcctcgtcaggtgcaggtcag 24
|||| | ||||| ||||||||||
Db 83 GATCTTGGTCAGCTGCAGGTCAG 61

RESULT 6
CNS026TZ
LOCUS CNS026TZ 872 bp DNA GSS 12-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 241J15 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL183824
VERSION AL183824.1 GI:7821928
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 872)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 872)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 872)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .872
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="241J15"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG241CE08SP1-end : PUC-Ori"
BASE COUNT 139 a 301 c 254 g 166 t 12 others
ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 872;
Best Local Similarity 87.0%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 cgatcctcgtcaggtgcaggtca 23
|| ||||| ||||||||||||
Db 715 CGCTCCTCTGCAGGTGCAGGTCA 737

RESULT 7
AL556337/c
LOCUS AL556337 883 bp mRNA EST 16-FEB-2001
DEFINITION AL556337 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK001YI14 5 prime, mRNA sequence.
ACCESSION AL556337
VERSION AL556337.1 GI:12898922
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 883)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source 1. .883
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK001YI14"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 193 a 237 c 228 g 224 t 1 others
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 883;
Best Local Similarity 87.0%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gatcctcgtcaggtgcaggtcag 24
|| ||||| ||||||| ||||||
Db 39 GAGCCTCGCCAGGTGCCGGTCAG 17

RESULT 8
BG023733/c
LOCUS BG023733 904 bp mRNA EST 24-JAN-2001
DEFINITION 602303347F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4395058 5', mRNA sequence.
ACCESSION BG023733
VERSION BG023733.1 GI:12408790
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 904)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

```
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM10091  row: m  column: 11
          High quality sequence stop: 733.
FEATURES             Location/Qualifiers
     source          1. .904
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4395058"
                     /clone_lib="NIH_MGC_88"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      190 a    253 c    239 g    222 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 11;  Length 904;
  Best Local Similarity   87.0%;  Pred. No. 1.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           || ||||| ||||||| ||||||
Db      33 GAGCCTCGCCAGGTGCCGGTCAG 11


RESULT   9
AL580709/c
LOCUS       AL580709      917 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL580709 LTI_NFL008_TC2 Homo sapiens cDNA clone CS0DJ008YM14 5
            prime, mRNA sequence.
ACCESSION   AL580709
VERSION     AL580709.1  GI:12946994
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 917)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .917
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ008YM14"
                     /clone_lib="LTI_NFL008_TC2"
                     /sex="male"
                     /tissue_type="T cells from T cell leukemia"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime  end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@liferech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      204 a    246 c    233 g    229 t      5 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 917;
  Best Local Similarity   87.0%;  Pred. No. 1.2e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           || ||||| ||||||| ||||||
Db      59 GAGCCTCGCCAGGTGCCGGTCAG 37


RESULT  10
BF979422/c
LOCUS       BF979422      1095 bp    mRNA            EST       23-JAN-2001
DEFINITION  602288251F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4373772 5',
            mRNA sequence.
ACCESSION   BF979422
VERSION     BF979422.1  GI:12346637
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1095)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10036  row: f  column: 13
            High quality sequence stop: 585.
FEATURES             Location/Qualifiers
     source          1. .1095
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4373772"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      248 a    314 c    343 g    188 t      2 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 11;  Length 1095;
  Best Local Similarity   87.0%;  Pred. No. 1.3e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           || ||||| ||||||| ||||||
Db     326 GAGCCTCGCCAGGTGCCGGTCAG 304


RESULT  11
AQ982558/c
LOCUS       AQ982558      494 bp    DNA             GSS       29-JAN-2000
DEFINITION  RPCI-23-294H19.TJ RPCI-23 Mus musculus genomic clone RPCI-23-294H19
            , DNA sequence.
ACCESSION   AQ982558
VERSION     AQ982558.1  GI:6814859
KEYWORDS    GSS.
```

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-294H19.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 294  row: H  column: 19
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .494
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="RPCI-23-294H19"
                      /clone_lib="RPCI-23"
                      /sex="Female"
                      /lab_host="DH10B"
                      /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                      EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                      brain genomic DNA was isolated and  partially digested
                      with a combination of EcoRI and EcoRI Methylase.   Size
                      selected DNA was cloned into the pBACe3.6 vector at the
                      EcoRI  sites.  The ligation products were transformed into
                      DH10B  electrocompetent cells (BRL Life Technologies).  "
BASE COUNT      167 a     94 c    130 g    102 t      1 others
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 13;  Length 494;
  Best Local Similarity   90.5%;  Pred. No. 1.7e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtca 23
           |||||||||| |||||| ||
Db      98 ATCCTCGTCAGTTGCAGGACA 78


RESULT  12
BE234470
LOCUS       BE234470      536 bp    mRNA            EST       10-JUL-2000
DEFINITION  141586 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BE234470
VERSION     BE234470.1  GI:9019188
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 79  row: F  column: 18
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES              Location/Qualifiers
     source           1. .536
                      /organism="Sus scrofa"
                      /db_xref="taxon:9823"
                      /clone_lib="MARC 1PIG"
                      /tissue_type="pooled"
                      /lab_host="DH10B"
                      /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                      Library made from pooled tissue from day 11, 13, 15, 20,
                      and 30 embryos."
BASE COUNT       92 a    126 c    170 g    146 t      2 others
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 10;  Length 536;
  Best Local Similarity   90.5%;  Pred. No. 1.7e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 atcctcgtcaggtgcaggtca 23
           ||||| | ||||||||||||||
Db     162 ATCCTGGGCAGGTGCAGGTCA 182


RESULT  13
AL538687
LOCUS       AL538687      777 bp    mRNA            EST       16-FEB-2001
DEFINITION  AL538687 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF037YP11 3
            prime, mRNA sequence.
ACCESSION   AL538687
VERSION     AL538687.1  GI:12867208
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 777)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES              Location/Qualifiers
     source           1. .777
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="CS0DF037YP11"
                      /clone_lib="LTI_FL013_FBrn1"
                      /dev_stage="pooled tissue from post conception fetuses (20
                      week, 24 week and 26 week)"
                      /lab_host="DH10B"
                      /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
                      cDNA was primed with a NotI-oligo(dT) primer. Five prime
                      end enriched, double-stranded cDNA was digested with Not I
                      and cloned into the Not I and Eco RV sites of the
                      pCMVSPORT 6 vector.  Library was constructed by Life
                      Technologies.  Contact : Feng Liang Life Technologies, a
```

```
                    division of Invitrogen 9800 Medical Center Drive Rockville
                    , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
                    fliang@lifetech.com URL :
                    http://fulllength.invitrogen.com"
BASE COUNT      173 a    220 c    205 g    169 t     10 others
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 10;  Length 777;
  Best Local Similarity   82.6%;  Pred. No. 1.8e+03;
  Matches   19;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcag 24
           | |||||||||| : ||||||||
Db     486 GGTCCTCGTCAGAKCCAGGTCAG 508


RESULT  14
CNS044UQ
LOCUS       CNS044UQ     1032 bp    DNA             GSS       18-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            082E16 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL274571
VERSION     AL274571.1  GI:8008771
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1032)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1032)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1032)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1032
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="082E16"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG082BC08LP1~end : T7"
BASE COUNT      267 a    249 c    220 g    261 t     35 others
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 13;  Length 1032;
  Best Local Similarity   90.5%;  Pred. No. 1.8e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 tcctcgtcaggtgcaggtcag 24
           |||||  ||||||||||||||
Db     753 TCCTCAGCAGGTGCAGGTCAG 773


RESULT  15
BI078629/c
LOCUS       BI078629     1064 bp    mRNA            EST       20-JUN-2001
DEFINITION  602872730F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5004798 5',
            mRNA sequence.
ACCESSION   BI078629
VERSION     BI078629.1  GI:14496959
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1064)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11043  row: k  column: 07
            High quality sequence start: 10
            High quality sequence stop: 418.
FEATURES             Location/Qualifiers
     source          1. .1064
                     /organism="Mus musculus"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5004798"
                     /clone_lib="NCI_CGAP_Mam2"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      274 a    332 c    283 g    175 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 11;  Length 1064;
  Best Local Similarity   90.5%;  Pred. No. 1.8e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 tcctcgtcaggtgcaggtcag 24
           || || ||||||||||||||||
Db     213 TCTTCCTCAGGTGCAGGTCAG 193


RESULT  16
AA594461
LOCUS       AA594461     339 bp    mRNA            EST       25-SEP-1997
DEFINITION  nl94c07.s1 NCI_CGAP_Co10 Homo sapiens cDNA clone IMAGE:1058316 3'
            similar to gb:M77198 RAC-BETA SERINE/THREONINE KINASE (HUMAN);,
            mRNA sequence.
ACCESSION   AA594461
VERSION     AA594461.1  GI:2409811
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
```

```
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.
            , Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 2573   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES              Location/Qualifiers
     source           1. .339
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1058316"
                      /clone_lib="NCI_CGAP_Co10"
                      /tissue_type="colon tumor RER+"
                      /lab_host="DH10B"
                      /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                      modified polylinker; 1st strand cDNA was prepared from
                      RER+ colon tumor, and was then primed with a Not 1 -
                      oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                      RI adaptors (Pharmacia), digested with Not I and cloned
                      into the Not I and Eco RI sites of the modified pT7T3
                      vector.  Library is normalized. Library was constructed by
                      Bento Soares and M. Fatima Bonaldo (N-Soares4). "
BASE COUNT       99 a     76 c     96 g     68 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 339;
  Best Local Similarity   83.3%;  Pred. No. 1.9e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcagc 25
           ||||| ||||| | ||| ||||||
Db      45 GATCCGCGTCACGCGCACGTCAGC 68


RESULT  17
BF318689/c
LOCUS       BF318689      360 bp    mRNA            EST       29-DEC-2000
DEFINITION  uy05a01.y1 McCarrey Eddy spermatocytes Mus musculus cDNA clone
            IMAGE:3657096 5' similar to TR:Q9VB28 Q9VB28 CG13972 PROTEIN. ;,
            mRNA sequence.
ACCESSION   BF318689
VERSION     BF318689.1  GI:11267302
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1419400
            High quality sequence stop: 338.
FEATURES              Location/Qualifiers
     source           1. .360
                      /organism="Mus musculus"
                      /strain="CD-1"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:3657096"
                      /clone_lib="McCarrey Eddy spermatocytes"
                      /sex="male"
                      /tissue_type="spermatocytes, pooled from multiple mice"
                      /dev_stage="60 day"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: testis; Vector: pBluescript SK+ (Stratagene
                      ); Site_1: XhoI; Site_2: EcoRI; cDNA oligo dT-primed
                      [5'-(GA)10-ACTAGTCTCGAGTTTTTTTTTTTTT-3'] and directionally
                      cloned  using 5' linkers 5'-AATTCGGCACGAG-3' and
                      5'-CTCGTGCCG-3'. Size selection of >400bp material gives
                      average insert size ranging from 1-2 kb. Library was mass
                      excised (from  lambda-UniZAP-XR) and resulting
                      single-stranded phagemids were prepped and tranformed
                      into DH10B. Library contains 98% recombinants.
                      References: J. Androl. 20:635-639 and Gene 25:263-269.
                      Library constructed and donated by J. McCarrey, Ph.D.
                      (Southwest Foundation for Biomedical Research, Dept. of
                      Genetics); excision done by E.M. Eddy, Ph.D. (National
                      Institutes of  Health, National Institute of Environmental
                      Health Sciences). Original lambda-based library is
                      available through ATCC, catalog #63422."
BASE COUNT      118 a     71 c     115 g     55 t      1 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 360;
  Best Local Similarity   80.0%;  Pred. No. 2e+03;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcagc 25
             |||| ||| |||| | |||||||| |
Db     181 CGATGCTCCTCAGCTNCAGGTCATC 157


RESULT  18
BF659032
LOCUS       BF659032      369 bp    mRNA            EST       20-DEC-2000
DEFINITION  maa89h11.y1 Soares_thymus_2NbMT Mus musculus cDNA clone
            IMAGE:3824132 5' similar to SW:PI4K_HUMAN P42356
            PHOSPHATIDYLINOSITOL 4-KINASE ALPHA ;, mRNA sequence.
ACCESSION   BF659032
VERSION     BF659032.1  GI:11924166
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1460484
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40RP from Gibco
            High quality sequence stop: 347.
FEATURES              Location/Qualifiers
     source           1. .369
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
```

/db_xref="taxon:10090"
/clone="IMAGE:3824132"
/clone_lib="Soares_thymus_2NbMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 70 a 121 c 104 g 74 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 11; Length 369;
Best Local Similarity 83.3%; Pred. No. 2e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 gatcctcgtcaggtgcaggtcagc 25
|| ||||||| |||||||| | ||
Db 52 GAGCCTCGTCCGGTGCAGGACTGC 75

RESULT 19
BE860402
LOCUS BE860402 398 bp mRNA EST 29-SEP-2000
DEFINITION UI-M-AN1-afc-d-12-0-UI.r1 NIH_BMAP_MBG_N Mus musculus cDNA clone UI-M-AN1-afc-d-12-0-UI 5', mRNA sequence.
ACCESSION BE860402
VERSION BE860402.1 GI:10377301
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 398)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .398
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AN1-afc-d-12-0-UI"
/clone_lib="NIH_BMAP_MBG_N"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_MBG_N library is a normalized library constructed from mouse basal ganglia. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories."

BASE COUNT 57 a 112 c 148 g 80 t 1 others
ORIGIN

Query Match 70.4%; Score 17.6; DB 11; Length 398;
Best Local Similarity 83.3%; Pred. No. 2e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 gatcctcgtcaggtgcaggtcagc 25
|| ||||||| |||||||| | ||
Db 78 GAGCCTCGTCCGGTGCAGGACTGC 101

RESULT 20
AA644735/c
LOCUS AA644735 413 bp mRNA EST 28-OCT-1997
DEFINITION M-EST008 Sugarcane mature stalk Saccharum sp. cDNA clone MA13-rev, mRNA sequence.
ACCESSION AA644735
VERSION AA644735.1 GI:2571144
KEYWORDS EST.
SOURCE Saccharum sp.
ORGANISM Saccharum sp.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE 1 (bases 1 to 413)
AUTHORS Carson,D.L. and Botha,F.C.
TITLE Sugarcane cDNA from mature stalk tissue
JOURNAL Unpublished (1997)
COMMENT Contact: Deborah Lee Carson
Biotechnology
South African Sugar Association Experiment Station
Private Bag X02, Mount Edgecombe, 4300, South Africa
Tel: 2731593205
Fax: 2731595406
Email: biotech@sugar.org.za
Partial sequence of a 600 bp clone
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .413
/organism="Saccharum sp."
/strain="variety Nco 376"
/db_xref="taxon:15819"
/clone="MA13-rev"
/clone_lib="Sugarcane mature stalk"
/note="Vector: Lambda ZAP II; Site_1: EcoRI; mRNA was prepared from mature stalk material of the commercial sugarcane variety Nco 376. cDNA was synthesized using standard protocols and ligated to an annealed amplification adaptor set. cDNA was then PCR amplified. Amplified cDNA was ligated and cloned into the Eco RI site of the Lambda ZAP II phage vector."

BASE COUNT 102 a 104 c 93 g 113 t 1 others
ORIGIN

Query Match 70.4%; Score 17.6; DB 10; Length 413;
Best Local Similarity 83.3%; Pred. No. 2e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 gatcctcgtcaggtgcaggtcagc 25
|||||| ||||| ||||||| ||
Db 188 GATCCTTGTCAGCTGCAGGTTTGC 165

```
RESULT  21
BE862745/c
LOCUS       BE862745      413 bp    mRNA            EST       29-SEP-2000
DEFINITION  UI-M-BH0-ajf-b-11-0-UI.r1 NIH_BMAP_M_S1 Mus musculus cDNA clone
            UI-M-BH0-ajf-b-11-0-UI 5', mRNA sequence.
ACCESSION   BE862745
VERSION     BE862745.1  GI:10382015
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH0-ajf-b-11-0-UI"
                     /clone_lib="NIH_BMAP_M_S1"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S1 library is a subtracted library derived from
                     a mixture of normalized libraries from ten regions of the
                     mouse brain (cerebellum, brain stems, olfactory bulbs,
                     hypothalamus, cortex, amygdala, basal ganglia, pineal
                     gland, striatum, hipoccampus). The driver used for
                     subtraction consisted of a pool of 20,000 cDNA clones
                     obtained from non-normalized and normalized libraries of
                     these ten regions of the mouse brain."
BASE COUNT       92 a    125 c    112 g     84 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 413;
  Best Local Similarity   83.3%;  Pred. No. 2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcag 24
           | ||   ||||||||||||||||
Db      24 CAATGTCCGTCAGGTGCAGGTCAC 1


RESULT  22
AQ172663/c
LOCUS       AQ172663      424 bp    DNA             GSS       17-OCT-1998
DEFINITION  HS_3194_A2_D09_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3194 Col=18 Row=G, DNA sequence.
ACCESSION   AQ172663
VERSION     AQ172663.1  GI:3570030
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3194  row: G  column: 18
            Class: BAC ends
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1. .424
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3194 Col=18 Row=G"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      119 a    110 c     94 g    101 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 424;
  Best Local Similarity   83.3%;  Pred. No. 2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 cgatcctcgtcaggtgcaggtcag 24
           | ||||| | | ||||||| |||||||
Db     382 CGATCTTGGCCAGGTGGAGGTCAG 359


RESULT  23
AQ371411/c
LOCUS       AQ371411      429 bp    DNA             GSS       06-MAR-1999
DEFINITION  HS_5043_A1_D05_SP6E RPCI11 Human Male BAC Library Homo sapiens
            genomic clone Plate=619 Col=9 Row=G, DNA sequence.
ACCESSION   AQ371411
VERSION     AQ371411.1  GI:4340516
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
```

```
           Plate: 619  row: G  column: 9
           Seq primer: SP6
           Class: BAC ends
           High quality sequence stop: 429.
FEATURES             Location/Qualifiers
     source          1. .429
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=619 Col=9 Row=G"
                     /clone_lib="RPCI11 Human Male BAC Library"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; RPCI11 Human Male BAC Library"
BASE COUNT      122 a     90 c    109 g    107 t      1 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 429;
  Best Local Similarity   83.3%;  Pred. No. 2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          || ||||| || || |||||||||
Db     52 GAGCCTCGGCAAGTTCAGGTCAGC 29


RESULT  24
BE850719
LOCUS       BE850719      463 bp    mRNA            EST       26-SEP-2000
DEFINITION  ux01a08.y1 Soares_thymus_2NbMT Mus musculus cDNA clone
            IMAGE:3470198 5', mRNA sequence.
ACCESSION   BE850719
VERSION     BE850719.1  GI:10309058
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1389558
            Seq primer: -40RP from Gibco
            High quality sequence stop: 451.
FEATURES             Location/Qualifiers
     source          1. .463
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3470198"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       85 a    123 c    173 g     82 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 463;
  Best Local Similarity   83.3%;  Pred. No. 2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          || ||||||| |||||||| | ||
Db    233 GAGCCTCGTCCGGTGCAGGACTGC 256


RESULT  25
AI789420/c
LOCUS       AI789420      495 bp    mRNA            EST       02-JUL-1999
DEFINITION  uk99a06.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2064658 5', mRNA sequence.
ACCESSION   AI789420
VERSION     AI789420.1  GI:5337136
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: uk99a06.x1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:993846
            Seq primer: custom primer used
            High quality sequence stop: 493.
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2064658"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      147 a    122 c    141 g     85 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 495;
  Best Local Similarity   83.3%;  Pred. No. 2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy      1 cgatcctcgtcaggtgcaggtcag 24
          | ||   ||||||||||||||||||
Db    400 CAATGTCCGTCAGGTGCAGGTCAG 377


RESULT  26
AI180705
LOCUS       AI180705      508 bp    mRNA            EST       08-OCT-1998
DEFINITION  ub91b01.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:1395817 5' similar to TR:Q13288 Q13288 P37NB. ;, mRNA
            sequence.
ACCESSION   AI180705
VERSION     AI180705.1  GI:3731343
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 508)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGT:907533
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .508
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1395817"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                     ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      112 a    135 c    146 g    114 t      1 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 508;
  Best Local Similarity   83.3%;  Pred. No. 2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcag 24
          |||||||||  | |  |||||| |||
Db     39 CGATCCTCGCCTGCTGCAGGACAG 62


RESULT  27
AU204314
LOCUS       AU204314      574 bp    mRNA            EST       17-JUL-2001
DEFINITION  AU204314 unpublished oligo-capped cDNA library, stage L1
            Caenorhabditis elegans cDNA clone yk832a03 5', mRNA sequence.
ACCESSION   AU204314
VERSION     AU204314.1  GI:14835617
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
            and Sugano,S.
  TITLE     A complementary view of the C.elegans genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yuji Kohara
            Genome Biology Lab.
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .574
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk832a03"
                     /clone_lib="unpublished oligo-capped cDNA library, stage
                     L1"
                     /sex="Hermaphrodite"
                     /tissue_type="whole animal"
                     /dev_stage="L1"
BASE COUNT      170 a    135 c    147 g    122 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 574;
  Best Local Similarity   83.3%;  Pred. No. 2.1e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 cgatcctcgtcaggtgcaggtcag 24
          | |||||||||  | ||||  |||||
Db    411 CCATCCTCGTCGGCTGCAAGTCAG 434


RESULT  28
AA536738
LOCUS       AA536738      618 bp    mRNA            EST       29-JUL-1997
DEFINITION  vj88c01.r1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
            IMAGE:944160 3' similar to SW:PI4K_HUMAN P42356
            PHOSPHATIDYLINOSITOL 4-KINASE ALPHA ;, mRNA sequence.
ACCESSION   AA536738
VERSION     AA536738.1  GI:2282731
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 618)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
```

```
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:541016
            Possible reversed clone: similarity on wrong strand
            High quality sequence stop: 443.
FEATURES             Location/Qualifiers
     source          1. .618
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:944160"
                     /clone_lib="Knowles Solter mouse 2 cell"
                     /tissue_type="embryo"
                     /dev_stage="2-cell"
                     /lab_host="DH10B"
                     /note="Organ: embryo; Vector: pBluescribe (modified);
                     Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
                     mRNA prepared from 13,500 2-cell stage embryos. Primer:
                     SalI(dT): 5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTT-3'. cDNAs
                     were cloned into the MluI/SalI sites of a modified
                     pBluescribe vector using commercial linkers (NEB).
                     Average insert size: 1.2 kb."
BASE COUNT      149 a    148 c    224 g     95 t      2 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 618;
  Best Local Similarity   83.3%;  Pred. No. 2.1e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          || ||||||| |||||||| | ||
Db    261 GAGCCTCGTCCGGTGCAGGACTGC 284


RESULT  29
BE653574
LOCUS       BE653574      628 bp    mRNA            EST       06-SEP-2000
DEFINITION  UI-M-AM1-agb-h-08-0-UI.r1 NIH_BMAP_MAM_N Mus musculus cDNA clone
            UI-M-AM1-agb-h-08-0-UI 5', mRNA sequence.
ACCESSION   BE653574
VERSION     BE653574.1  GI:9979482
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-AM1-agb-h-08-0-UI"
                     /clone_lib="NIH_BMAP_MAM_N"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_MAM_N library is a normalized library constructed
                     from mouse amygdala. The tag is a string of 5 nucleotides
                     present between the Not I site and the oligo-dT track.
                     The library was constructed as described by Bonaldo,
                     Lennon and Soares, Genome Research 6: 791-806, 1996.
                     Tissue provided by Ms. Annie Novakovich, Zivic-Miller
                     Laboratories."
BASE COUNT      121 a    172 c    207 g    128 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 628;
  Best Local Similarity   83.3%;  Pred. No. 2.1e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          || ||||||| |||||||| | ||
Db    430 GAGCCTCGTCCGGTGCAGGACTGC 453


RESULT  30
AZ426129/c
LOCUS       AZ426129      641 bp    DNA             GSS       03-OCT-2000
DEFINITION  1M0206H03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0206H03 R, DNA sequence.
ACCESSION   AZ426129
VERSION     AZ426129.1  GI:10550142
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0206  row: H  column: 03
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 641.
FEATURES             Location/Qualifiers
     source          1. .641
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0206H03"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained from the Jackson
                     Laboratory Mouse DNA Resource
```

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 156 a 184 c 183 g 118 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 13; Length 641;
Best Local Similarity 83.3%; Pred. No. 2.1e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cgatcctcgtcaggtgcaggtcag 24
| || |||||||||||||||||
Db 219 CAATGTCCGTCAGGTGCAGGTCAG 196

RESULT 31
BE262905/c
LOCUS BE262905 721 bp mRNA EST 26-OCT-2000
DEFINITION 601153594F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3509542 5', mRNA sequence.
ACCESSION BE262905
VERSION BE262905.1 GI:9136326
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 721)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM192 row: l column: 23
High quality sequence start: 56
High quality sequence stop: 575.
FEATURES Location/Qualifiers
source 1. .721
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3509542"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 186 a 216 c 200 g 119 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 10; Length 721;
Best Local Similarity 83.3%; Pred. No. 2.1e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 gatcctcgtcaggtgcaggtcagc 25
| ||||||| ||||||||| |||
Db 338 GGTCCTCGTGTGGTGCAGGTAAGC 315

RESULT 32
BG478947/c
LOCUS BG478947 833 bp mRNA EST 21-MAR-2001
DEFINITION 602525709F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4649354 5', mRNA sequence.
ACCESSION BG478947
VERSION BG478947.1 GI:13411226
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 833)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLCM1429 row: a column: 03
High quality sequence stop: 622.
FEATURES Location/Qualifiers
source 1. .833
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4649354"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 178 a 291 c 263 g 101 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 11; Length 833;
Best Local Similarity 83.3%; Pred. No. 2.1e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cgatcctcgtcaggtgcaggtcag 24
|| |||||||| ||||| |||| |
Db 572 CGGTCCTCGTCTGGTGCTGGTCCG 549

RESULT 33
BF794272/c
LOCUS BF794272 893 bp mRNA EST 12-JAN-2001
DEFINITION 602255473F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4338715 5', mRNA sequence.

```
ACCESSION   BF794272
VERSION     BF794272.1  GI:12099326
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 893)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis Staudt, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9948  row: a  column: 20
            High quality sequence stop: 672.
FEATURES             Location/Qualifiers
     source          1. .893
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4338715"
                     /clone_lib="NIH_MGC_85"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      204 a    229 c    264 g    196 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 893;
  Best Local Similarity   83.3%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcagc 25
           | ||||| ||||| ||||| ||||
Db     584 GCTCCTCCTCAGGAGCAGGGCAGC 561


RESULT  34
CNS02FRJ
LOCUS       CNS02FRJ      897 bp    DNA             GSS       13-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            135O01 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL195400
VERSION     AL195400.1  GI:7833506
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 897)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 897)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .897
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="135O01"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG135AH01LP1~end : T7"
BASE COUNT      185 a    227 c    259 g    223 t      3 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 897;
  Best Local Similarity   83.3%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcagc 25
           | ||||| ||||| ||||| ||||
Db     808 GATCCTCTGCAGGTTCAGGTGAGC 831


RESULT  35
BF313136/c
LOCUS       BF313136      929 bp    mRNA            EST       21-NOV-2000
DEFINITION  601897654F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4127005 5',
            mRNA sequence.
ACCESSION   BF313136
VERSION     BF313136.1  GI:11261079
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 929)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1017  row: l  column: 14
            High quality sequence stop: 615.
FEATURES             Location/Qualifiers
     source          1. .929
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4127005"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
```

```
                    Note: this is a NIH_MGC Library."
BASE COUNT      232 a    214 c    280 g    203 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 929;
  Best Local Similarity   83.3%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          | ||||| ||||| ||||| ||||
Db    599 GCTCCTCCTCAGGAGCAGGGCAGC 576


RESULT  36
CNS00L1P
LOCUS       CNS00L1P      942 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR23G17 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL067857
VERSION     AL067857.1  GI:4957886
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 942)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .942
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR23G17"
                     /note="end : T7"
BASE COUNT      292 a    166 c    183 g    242 t     59 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 942;
  Best Local Similarity   83.3%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          ||| || |||| |||||| |||||
Db    191 GATTCTGGTCATGTGCAGATCAGC 214


RESULT  37
CNS006PE/c
LOCUS       CNS006PE      982 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14J05 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065656
VERSION     AL065656.1  GI:4944725
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 982)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .982
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR14J05"
                     /note="end : T7"
BASE COUNT      205 a    289 c    287 g    161 t     40 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 982;
  Best Local Similarity   83.3%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 gatcctcgtcaggtgcaggtcagc 25
          ||||| |||| || || |||||||
Db    116 GATCCACGTCGGGAGCGGGTCAGC 93


RESULT  38
CNS00HWP/c
LOCUS       CNS00HWP      985 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR35J21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL073836
VERSION     AL073836.1  GI:4953810
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 985)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
```

```
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .985
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR35J21"
                     /note="end : TET3"
BASE COUNT      204 a    275 c    300 g    164 t     42 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 985;
  Best Local Similarity   83.3%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcagc 25
           ||||| |||| || || |||||||
Db     117 GATCCACGTCGGGAGCGGGTCAGC 94


RESULT  39
BF965985/c
LOCUS       BF965985     1002 bp    mRNA            EST       22-JAN-2001
DEFINITION  602277168F2 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4364968 5',
            mRNA sequence.
ACCESSION   BF965985
VERSION     BF965985.1  GI:12333200
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1002)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10013  row: g  column: 17
            High quality sequence stop: 570.
FEATURES             Location/Qualifiers
     source          1. .1002
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4364968"
                     /clone_lib="NIH_MGC_86"
                     /tissue_type="osteosarcoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bone; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.533 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      266 a    266 c    288 g    182 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 1002;
  Best Local Similarity   83.3%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcagc 25
           | ||||| ||||| ||||| ||||
Db     649 GCTCCTCCTCAGGAGCAGGGCAGC 626


RESULT  40
CNS00HQI/c
LOCUS       CNS00HQI     1016 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR35L24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL074118
VERSION     AL074118.1  GI:4953595
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1016)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1016
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR35L24"
                     /note="end : TET3"
BASE COUNT      201 a    295 c    299 g    169 t     52 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 1016;
  Best Local Similarity   83.3%;  Pred. No. 2.2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcaggtcagc 25
           ||||| |||| || || |||||||
Db     118 GATCCACGTCGGGAGCGGGTCAGC 95


RESULT  41
BE795069/c
LOCUS       BE795069     1126 bp    mRNA            EST       20-SEP-2000
DEFINITION  601592885F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3946869 5',
            mRNA sequence.
```

```
ACCESSION   BE795069
VERSION     BE795069.1  GI:10216267
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1126)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM807  row: j  column: 22
            High quality sequence start: 10
            High quality sequence stop: 702.
FEATURES             Location/Qualifiers
     source          1. .1126
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3946869"
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      262 a    320 c    289 g    255 t
ORIGIN
```

```
  Query Match             69.6%;  Score 17.4;  DB 11;  Length 1126;
  Best Local Similarity   94.7%;  Pred. No. 2.7e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 gatcctcgtcaggtgcagg 20
           |||||||||||||||| |||
Db     853 GATCCTCGTCAGGTGTAGG 835
```

```
RESULT  42
BG328084/c
LOCUS       BG328084     1298 bp    mRNA            EST       27-FEB-2001
DEFINITION  602427134F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546497 5',
            mRNA sequence.
ACCESSION   BG328084
VERSION     BG328084.1  GI:13134522
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1298)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1233  row: c  column: 10
            High quality sequence stop: 511.
FEATURES             Location/Qualifiers
     source          1. .1298
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4546497"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      316 a    498 c    357 g    127 t
ORIGIN
```

```
  Query Match             69.6%;  Score 17.4;  DB 11;  Length 1298;
  Best Local Similarity   94.7%;  Pred. No. 2.7e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 tcctcgtcaggtgcaggtc 22
           || ||||||||||||||||
Db     987 TCGTCGTCAGGTGCAGGTC 969
```

```
RESULT  43
AW371554
LOCUS       AW371554     108 bp    mRNA            EST       04-FEB-2000
DEFINITION  QV3-BT0295-111199-017-g08 BT0295 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW371554
VERSION     AW371554.1  GI:6876208
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 108)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV3&t2=QV3-BT0295-
            111199-017-g08&t3=1999-11-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 108.
FEATURES             Location/Qualifiers
     source          1. .108
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0295"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
```

```
                ,716 -  Ludwig Institute for Cancer Research) profiles
                into  the pUC 18 vector. Reverse transcription of tissue
                mRNA and cDNA amplification were performed under low
                stringency conditions."
BASE COUNT        8 a     39 c     23 g     38 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 10;  Length 108;
  Best Local Similarity   86.4%;  Pred. No. 2.5e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcctcgtcaggtgcaggtcagc 25
          ||||| ||||| ||||| ||||
Db      1 TCCTCCTCAGGAGCAGGGCAGC 22


RESULT  44
AA651689/c
LOCUS       AA651689      112 bp    mRNA            EST       13-NOV-1997
DEFINITION  nn47b06.r1 NCI_CGAP_Kid6 Homo sapiens cDNA clone IMAGE:1086995 5'
            similar to gb:X63526 ELONGATION FACTOR 1-GAMMA (HUMAN);, mRNA
            sequence.
ACCESSION   AA651689
VERSION     AA651689.1  GI:2583341
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Stratagene, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 815   Std Error: 0.00
            Seq primer: -28m13 rev1 ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .112
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1086995"
                      /clone_lib="NCI_CGAP_Kid6"
                      /sex="mixed"
                      /tissue_type="kidney tumor"
                      /lab_host="SOLR (kanamycin resistant)"
                      /note="Organ: kidney; Vector: Bluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. Pooled kidney tumors. 5' adaptor sequence: 5'
                      GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTTT 3' Average insert size:  1.0 kb."
BASE COUNT       44 a     21 c     36 g     11 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 10;  Length 112;
  Best Local Similarity   86.4%;  Pred. No. 2.5e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcctcgtcaggtgcaggtcagc 25
          ||||| ||||| ||||| ||||
Db    101 TCCTCCTCAGGAGCAGGGCAGC 80


RESULT  45
AA197041/c
LOCUS       AA197041      235 bp    mRNA            EST       12-MAR-1998
DEFINITION  zq10c11.r1 Stratagene muscle 937209 Homo sapiens cDNA clone
            IMAGE:629300 5' similar to gb:X63526 ELONGATION FACTOR 1-GAMMA
            (HUMAN);, mRNA sequence.
ACCESSION   AA197041
VERSION     AA197041.1  GI:1792665
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 235)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 757   Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 138.
FEATURES              Location/Qualifiers
     source           1. .235
                      /organism="Homo sapiens"
                      /db_xref="GDB:5049534"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:629300"
                      /clone_lib="Stratagene muscle 937209"
                      /tissue_type="muscle"
                      /dev_stage="adult"
                      /lab_host="SOLR (kanamycin resistant)"
                      /note="Organ: skeletal muscle; Vector: pBluescript SK-;
                      Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally.
                      Primer: Oligo dT. Skeletal muscle from patient with
                      malignant hyperthermia.  Average insert size: 1.0 kb;
                      Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
                      3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT       56 a     52 c     71 g     45 t     11 others
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 10;  Length 235;
  Best Local Similarity   86.4%;  Pred. No. 2.7e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcctcgtcaggtgcaggtcagc 25
          ||||| ||||| ||||| ||||
Db     83 TCCTCCTCAGGAGCAGGGCAGC 62


Search completed: December 26, 2001, 12:12:20
Job time: 6502 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:58:48 ; Search time 1755.64 Seconds
(without alignments)
234.917 Million cell updates/sec

Title: US-09-396-196F-10
Perfect score: 25
Sequence: 1 tcgtcaggtgcaggtcagcacgttg 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		25	100.0	1041	6	AR029499	AR029499 Sequence
2		25	100.0	1041	6	AR034916	AR034916 Sequence
3		25	100.0	1084	6	A11530	A11530 BioB gene o
4		25	100.0	1121	6	E00893	E00893 Genomic DNA
5		25	100.0	5793	1	ECOBIO	J04423 E.coli 7,8-
6		25	100.0	5872	6	A38246	A38246 Sequence 1
7		25	100.0	5872	6	A38251	A38251 Sequence 6
8		25	100.0	5872	6	A93674	A93674 Sequence 1
9		25	100.0	5872	6	A93679	A93679 Sequence 6
10		25	100.0	5872	6	AR101809	AR101809 Sequence
11		25	100.0	5872	6	AR101810	AR101810 Sequence
12		25	100.0	11022	1	AE000180	AE000180 Escherich
13		23.4	93.6	5526	1	AF250776	AF250776 Unculture
14		23.4	93.6	13501	1	AE005258	AE005258 Escherich
15		23.4	93.6	297816	1	AP002553	AP002553 Escherich
16		22.4	89.6	7215	1	SMABIO	D17468 S. marcesce
17		20.8	83.2	4000	1	AF250768	AF250768 Unculture
18		20.2	80.8	1041	1	EHBIOTOP2	U38648 Erwinia her
19	c	19.4	77.6	213732	1	AE001862	AE001862 Deinococc
20	c	18.6	74.4	50195	2	AC019891	AC019891 Drosophil
21		18.6	74.4	73076	3	AC003055	AC003055 Drosophil
22	c	18.6	74.4	168540	3	AC092189	AC092189 Drosophil
23		18.6	74.4	175369	2	AC026904	AC026904 Homo sapi
24	c	18.6	74.4	295312	3	AE003582	AE003582 Drosophil
25		18.2	72.8	72289	9	HSDJ911I5	AL050316 Human DNA
26		18.2	72.8	72919	8	AC003096	AC003096 Arabidops
27		18.2	72.8	83122	8	ATT20O10	AL163816 Arabidops
28	c	18.2	72.8	93330	2	AC022591	AC022591 Homo sapi
29	c	18.2	72.8	122951	9	AF312913	AF312913 Homo sapi
30	c	18.2	72.8	141990	9	AC004691	AC004691 Homo sapi
31		18.2	72.8	180257	2	AC007044	AC007044 Homo sapi
32		18.2	72.8	180484	2	AC091668	AC091668 Pan trogl
33	c	18.2	72.8	184042	2	AC092408	AC092408 Papio cyn
34		18.2	72.8	202828	33	AC064835	Ac064835 Homo sapi
35	c	18.2	72.8	294800	1	SME591789	AL591789 Sinorhizo
36		17.8	71.2	12891	1	AE004192	AE004192 Vibrio ch
37	c	17.6	70.4	694	1	AF145866	AF145866 Metal-con
38	c	17.6	70.4	1412	1	AF269004	AF269004 Unculture
39	c	17.6	70.4	4877	9	HUMHXBA	M24630 Human hexab
40	c	17.6	70.4	7286	9	HSTNSCN	X56160 Human mRNA
41	c	17.6	70.4	7346	6	AX022893	AX022893 Sequence
42	c	17.6	70.4	7346	6	AX030481	AX030481 Sequence
43	c	17.6	70.4	7390	9	HUMHXB	M55618 Homo sapien
44	c	17.6	70.4	7560	6	AX045676	AX045676 Sequence
45	c	17.6	70.4	7560	9	HSTENAS3	X78565 H.sapiens m
46		17.6	70.4	11213	1	AE004950	AE004950 Pseudomon
47	c	17.6	70.4	11606	1	AE006963	AE006963 Mycobacte
48	c	17.6	70.4	19770	1	MTCI376	Z95972 Mycobacteri
49	c	17.6	70.4	53533	1	BSY09476	Y09476 B.subtilis
50	c	17.6	70.4	72360	2	AC022984	AC022984 Homo sapi
51		17.6	70.4	104294	2	RN510D20	AL603723 Rattus no
52		17.6	70.4	143611	9	AC068992	AC068992 Homo sapi
53		17.6	70.4	144483	2	AC091701	AC091701 Trypanoso
54		17.6	70.4	159151	2	AC087347	AC087347 Homo sapi
55		17.6	70.4	169172	2	AC026663	AC026663 Homo sapi
56	c	17.6	70.4	180238	2	AC073688	AC073688 Mus muscu
57	c	17.6	70.4	191320	2	AC012568	AC012568 Homo sapi
58		17.6	70.4	200000	9	AP000494	AP000494 Homo sapi
59	c	17.6	70.4	202598	2	AC009813	AC009813 Homo sapi
60	c	17.6	70.4	210440	1	BSUB0006	Z99109 Bacillus su
61		17.6	70.4	232190	2	AC074167	AC074167 Mus muscu
62	c	17.4	69.6	629	9	HS10PT056	Z96156 H.sapiens t
63	c	17.4	69.6	1218	10	AF044602	AF044602 Mus muscu
64	c	17.4	69.6	10550	9	HUMIL9RPS	L39063 Homo sapien
65	c	17.4	69.6	127070	9	AC073052	AC073052 Homo sapi
66		17.4	69.6	174791	2	AC022391	AC022391 Homo sapi
67	c	17.4	69.6	182396	2	AC092679	AC092679 Homo sapi
68	c	17.4	69.6	185492	33	AC025339	Ac025339 Homo sapi
69		17.2	68.8	1135	3	AF230347	AF230347 Trichonym
70		17.2	68.8	1135	3	AF230348	AF230348 Trichonym

	71	17.2	68.8	1671	8	AF008121 AF008121 Eleusine
c	72	17.2	68.8	10944	1	AE002563 AE002563 Neisseria
	73	17.2	68.8	38065	1	AL035159 AL035159 Mycobacte
c	74	17.2	68.8	64110	9	AC025570 AC025570 Homo sapi
c	75	17.2	68.8	111301	8	AC011717 AC011717 Arabidops
c	76	17.2	68.8	135116	2	AP002437 AP002437 Homo sapi
c	77	17.2	68.8	167386	2	AC013704 AC013704 Homo sapi
c	78	17.2	68.8	172325	6	AX044035 AX044035 Sequence
c	79	17.2	68.8	178420	9	AP003548 AP003548 Homo sapi
c	80	17.2	68.8	179110	9	AP003477 AP003477 Homo sapi
	81	17.2	68.8	185703	2	AC026233 AC026233 Mus muscu
	82	17.2	68.8	193044	2	AC027769 AC027769 Homo sapi
	83	17.2	68.8	195516	2	AC053540 AC053540 Homo sapi
	84	17.2	68.8	199569	9	AC016250 AC016250 Homo sapi
	85	17.2	68.8	203802	2	AC016725 AC016725 Homo sapi
	86	17.2	68.8	211076	2	AC012214 AC012214 Homo sapi
c	87	17.2	68.8	211118	2	AC073752 AC073752 Mus muscu
	88	17.2	68.8	346294	1	AP002999 AP002999 Mesorhizo
	89	17	68.0	1155	6	AX121143 AX121143 Sequence
c	90	17	68.0	1572	6	AX120086 AX120086 Sequence
c	91	17	68.0	1851	10	MMU77683 U77683 Mus musculu
c	92	17	68.0	1881	10	MMU77684 U77684 Mus musculu
c	93	17	68.0	1883	10	MUSRXRA M84817 Mouse retin
	94	17	68.0	1934	9	HOSAWSCR5 AF041059 Homo sapi
	95	17	68.0	2081	10	HAMHPP M96650 Mesocricetu
c	96	17	68.0	2088	10	MMRXRAMR X66223 M.musculus
c	97	17	68.0	2095	6	I90008 I90008 Sequence 3
	98	17	68.0	2877	10	MMGNA11S03 U37413 Mus musculu
c	99	17	68.0	3521	22	E11161 E11161 Genomic DNA
c	100	17	68.0	7318	10	AB026497 AB026497 Mus muscu

ALIGNMENTS

```
RESULT   1
AR029499
LOCUS       AR029499     1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1  GI:5941472
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.Andrew.
  TITLE     Enhanced biotin biosynthesis in plant tissue
  JOURNAL   Patent: US 5859335-A 7 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.55;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    111 TCGTCAGGTGCAGGTCAGCACGTTG 135


RESULT   2
AR034916
LOCUS       AR034916     1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1  GI:5950521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.A.
  TITLE     Transgenic plants having increased biotin content
  JOURNAL   Patent: US 5869719-A 7 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.55;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    111 TCGTCAGGTGCAGGTCAGCACGTTG 135


RESULT   3
A11530
LOCUS       A11530       1084 bp    DNA             PAT       09-FEB-1994
DEFINITION  BioB gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1  GI:490218
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   .
  JOURNAL   Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            24. .1064
                     /gene="BIOB"
     CDS             24. .1064
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490219"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
BASE COUNT      271 a    286 c    318 g    209 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1084;
  Best Local Similarity  100.0%;  Pred. No. 0.54;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    134 TCGTCAGGTGCAGGTCAGCACGTTG 158


RESULT   4
E00893
LOCUS       E00893       1121 bp    DNA             PAT       29-SEP-1997
DEFINITION  Genomic DNA encoding biotin Synthetase.
```

```
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES              Location/Qualifiers
     source           1. .1121
                      /organism="Escherichia coli"
                      /db_xref="taxon:562"
BASE COUNT      289 a    296 c    325 g    211 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1121;
  Best Local Similarity  100.0%;  Pred. No. 0.54;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    152 TCGTCAGGTGCAGGTCAGCACGTTG 176


RESULT   5
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES              Location/Qualifiers
     source           1. .5793
                      /organism="Escherichia coli"
                      /strain="K-12"
                      /db_xref="taxon:562"
     CDS              complement(98. .574)
                      /note="ORF 1"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="AAA23513.1"
                      /db_xref="GI:455168"
                      /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                      SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                      YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene             complement(633. .1925)
                      /gene="bioA"
     CDS              complement(633. .1925)
                      /gene="bioA"
                      /codon_start=1
                      /transl_table=11
                      /product="7,8-diamino-pelargonic acid aminotransferase"
                      /protein_id="AAA23514.1"
                      /db_xref="GI:457106"
                      /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                      LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                      ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                      SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                      GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                      LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                      QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                      PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene             2012. .3052
                      /gene="bioB"
     CDS              2012. .3052
                      /gene="bioB"
                      /codon_start=1
                      /transl_table=11
                      /product="biotin synthetase"
                      /protein_id="AAA23515.1"
                      /db_xref="GI:145425"
                      /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                      LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                      KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                      YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                      VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                      FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                      PDTDEYYNAAAL"
     gene             3049. .4203
                      /gene="bioF"
     CDS              3049. .4203
                      /gene="bioF"
                      /codon_start=1
                      /transl_table=11
                      /product="7-keto-8-amino pelargonic acid synthetase"
                      /protein_id="AAA23516.1"
                      /db_xref="GI:145426"
                      /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                      SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                      LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                      LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                      SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                      RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                      AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene             4190. .4945
                      /gene="bioC"
     CDS              4190. .4945
                      /gene="bioC"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="AAA23517.1"
                      /db_xref="GI:145427"
                      /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                      HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                      TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
```

```
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4938. .5597
                     /gene="bioD"
     CDS             4938. .5597
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAA23518.1"
                     /db_xref="GI:145428"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                     T"
BASE COUNT     1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match             100.0%;  Score 25;  DB 1;  Length 5793;
  Best Local Similarity   100.0%;  Pred. No. 0.43;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db   2122 TCGTCAGGTGCAGGTCAGCACGTTG 2146


RESULT   6
A38246
LOCUS       A38246       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /number=1
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAA02325.1"
                     /db_xref="GI:2294845"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=3
                     /product="PROTEIN"
                     /protein_id="CAA02326.1"
                     /db_xref="GI:2294846"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:
                     8-AMINO-7-OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /number=5
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAA02327.1"
                     /db_xref="GI:2294847"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=6
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAA02328.1"
                     /db_xref="GI:2294848"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
```

```
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.43;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT   7
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.43;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT   8
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
```

```
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN
```

```
  Query Match            100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity  100.0%;  Pred. No. 0.43;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     227 TCGTCAGGTGCAGGTCAGCACGTTG 251
```

```
RESULT   9
A93679
LOCUS       A93679       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN
```

```
  Query Match            100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity  100.0%;  Pred. No. 0.43;
```

```
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT  10
AR101809
LOCUS       AR101809     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 1 from patent US 6083712.
ACCESSION   AR101809
VERSION     AR101809.1  GI:12812607
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.43;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT  11
AR101810
LOCUS       AR101810     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 6 from patent US 6083712.
ACCESSION   AR101810
VERSION     AR101810.1  GI:12812608
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.43;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT  12
AE000180
LOCUS       AE000180    11022 bp    DNA             BCT       01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 70 of 400 of the complete
            genome.
ACCESSION   AE000180 U00096
VERSION     AE000180.1  GI:1786988
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 11022)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 11022)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data.  Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .11022
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   <1. .130
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 4 REP sequences"
     gene            complement(147. .1430)
                     /gene="ybhC"
                     /note="b0772"
```

```
CDS             complement(147. .1430)
                /gene="ybhC"
                /function="orf; Not classified"
                /note="f427; 98 pct identical to fragment YBHC_ECOLI
                SW:P46130 (300 aa) but contains 127 additional C-terminal
                residues"
                /codon_start=1
                /transl_table=11
                /product="putative pectinesterase"
                /protein_id="AAC73859.1"
                /db_xref="GI:1786989"
                /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPAGTQGVTHTTIQAAVD
                AAIIKRTNKRQYIAVMPGEYQGTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
                NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSNIYYGFLAVNSRFN
                AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
promoter        complement(1507. .1535)
                /note="factor Sigma70; predicted +1 start at 806574"
gene            complement(1582. .2058)
                /gene="ybhB"
                /note="b0773"
CDS             complement(1582. .2058)
                /gene="ybhB"
                /function="orf; Unknown"
                /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC73860.1"
                /db_xref="GI:1786990"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
protein_bind    complement(2108. .2124)
                /note="central position to predicted promoter:85.5"
                /bound_moiety="RhaS predicted site"
gene            complement(2117. .3406)
                /gene="bioA"
                /note="b0774"
CDS             complement(2117. .3406)
                /gene="bioA"
                /EC_number="2.6.1.62"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                /codon_start=1
                /transl_table=11
                /product="7,8-diaminopelargonic acid synthetase"
                /protein_id="AAC73861.1"
                /db_xref="GI:1786991"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
promoter        complement(2193. .2221)
                /gene="bioA"
                /note="factor Sigma70; predicted +1 start at 807260"
protein_bind    3411. .3450
                /note="central position to bioB promoter: -20"
                /bound_moiety="BioB documented site"
protein_bind    3411. .3450
                /note="central position to predicted promoter: -20;genetic
                evidence for the site"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to predicted promoter:50"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to bioA promoter:50"
                /bound_moiety="BioB documented site"
promoter        3413. .3441
                /note="factor Sigma70; promoter bioB; documented +1 at
                808525"
promoter        complement(3447. .3473)
                /note="factor Sigma70; promoter bioA; documented +1 at
                808515"
gene            3493. .4533
                /gene="bioB"
                /note="b0775"
CDS             3493. .4533
                /gene="bioB"
                /EC_number="2.8.1.-"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                /codon_start=1
                /transl_table=11
                /product="biotin synthesis, sulfur insertion?"
                /protein_id="AAC73862.1"
                /db_xref="GI:1786992"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            4530. .5684
                /gene="bioF"
                /note="b0776"
CDS             4530. .5684
                /gene="bioF"
                /EC_number="2.3.1.47"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                /codon_start=1
                /transl_table=11
                /product="8-amino-7-oxononanoate synthase"
                /protein_id="AAC73863.1"
                /db_xref="GI:1786993"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
```

```
  Query Match           100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity  100.0%;  Pred. No. 0.39;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db   3603 TCGTCAGGTGCAGGTCAGCACGTTG 3627


RESULT  13
AF250776
LOCUS       AF250776     5526 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
            modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
            biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
            biotin biosynthesis protein BioC (bioC) genes, complete cds; and
            dethiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION   AF250776
VERSION     AF250776.1  GI:12620124
KEYWORDS    .
SOURCE      uncultured bacterium pCosHE2.
  ORGANISM  uncultured bacterium pCosHE2
```

```
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .5526
                     /organism="uncultured bacterium pCosHE2"
                     /db_xref="taxon:143797"
                     /clone="pCosHE2"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     CDS             complement(52. .528)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical 17.1 kDa protein in modC-bioA
                     intergenic region"
                     /protein_id="AAG60577.1"
                     /db_xref="GI:12620125"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKAG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(587. .1876)
                     /gene="bioA"
     CDS             complement(587. .1876)
                     /gene="bioA"
                     /note="7,8-diaminopelargonic acid
                     synthetase-aminotransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="DAPA-aminotransferase BioA"
                     /protein_id="AAG60578.1"
                     /db_xref="GI:12620126"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LKCVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            1963. .3003
                     /gene="bioB"
     CDS             1963. .3003
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase BioB"
                     /protein_id="AAG60579.1"
                     /db_xref="GI:12620127"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPNERDMPYLEQMVQGVKDLGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3000. .4154
                     /gene="bioF"
     CDS             3000. .4154
                     /gene="bioF"
                     /note="8-amino-7-oxononanoate synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="KAPA synthetase BioF"
                     /protein_id="AAG60580.1"
                     /db_xref="GI:12620128"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQLVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRRDEGDARREKLVSLIAHFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4141. .4896
                     /gene="bioC"
     CDS             4141. .4896
                     /gene="bioC"
                     /note="reaction step prior to synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin biosynthesis protein BioC"
                     /protein_id="AAG60581.1"
                     /db_xref="GI:12620129"
                     /translation="MATVNKQAIAAAFGRAAALYEQHADLQRQSADVLLAMLPERKYT
                     HVLDAGCGPGWMTRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRSKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4883. .>5526
                     /gene="bioD"
     CDS             4883. .>5526
                     /gene="bioD"
                     /note="DTB synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase BioD"
                     /protein_id="AAG60582.1"
                     /db_xref="GI:12620130"
                     /translation="MLVSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVAS
                     GSEKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESSVMS
                     SGLRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAML
                     TAQAIQHAGLTLAGWVANDVTPPGKRHAEYITTLTRMIPAPLLGEIPWLAENPE"
BASE COUNT     1274 a   1507 c   1567 g   1178 t
ORIGIN


  Query Match             93.6%;  Score 23.4;  DB 1;  Length 5526;
  Best Local Similarity   96.0%;  Pred. No. 2.3;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          ||||||||| ||||||||||||||||
Db   2073 TCGTCAGGTCCAGGTCAGCACGTTG 2097


RESULT  14
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
```

```
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .13501
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .7576
                     /note="O-island #36; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655; Cryptic prophage
                     CP-933K; includes one copy of the 13 bp direct repeat that
                     flanks the prophage"
     gene            66. .665
                     /gene="lomK"
                     /note="Z0981"
     CDS             66. .665
                     /gene="lomK"
                     /function="putative membrane; Other or unknown (Phage or
                     Prophage Related)"
                     /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                     residues 1 to 199 of 199 from GenPept 118 :
                     gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
                     [prophage P-EibA]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein of prophage
                     CP-933K"
                     /protein_id="AAG55137.1"
                     /db_xref="GI:12513752"
                     /translation="MRKVCAAILSAAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG
                     SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG
                     PSVRVNEWFSAYAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW
                     GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"
     gene            719. .2041
                     /gene="Z0982"
     CDS             719. .2041
                     /gene="Z0982"
                     /function="putative structure; Structural component (Phage
                     or Prophage Related)"
                     /note="Residues 164 to 440 of 440 are 68.79 pct identical
                     to residues 381 to 645 of 645 from GenPept 118 :
                     gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
                     tail fiber protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tail component of prophage CP-933K"
                     /protein_id="AAG55138.1"
                     /db_xref="GI:12513753"
                     /translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE
                     NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQPGTLNDFLGAMTEDDAR
                     PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE
                     SARQAAESAASAKKSEEASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA
                     TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE
                     RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA
                     GPQGPKGETGAAGPVGATGPQGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI
                     TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"
     gene            2004. .2312
                     /gene="Z0984"
     CDS             2004. .2312
                     /gene="Z0984"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 102 of 102 are 98.03 pct identical to
                     residues 28 to 129 of 129 from GenPept 118 :
                     gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
                     hypothetical protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55139.1"
                     /db_xref="GI:12513754"
                     /translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR
                     LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK
                     "
     gene            2489. .3469
                     /gene="Z0985"
     CDS             2489. .3469
                     /gene="Z0985"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 13 to 326 of 326 are 55.73 pct identical
                     to residues 18 to 331 of 336 from GenPept 118 :
                     gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
                     predicted by Glimmer [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55140.1"
                     /db_xref="GI:12513755"
                     /translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF
                     QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS
                     KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR
                     EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD
                     MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD
                     AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"
     gene            3530. .4522
                     /gene="Z0986"
     CDS             3530. .4522
                     /gene="Z0986"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55141.1"
                     /db_xref="GI:12513756"
                     /translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
                     SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK
                     VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY
                     VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG
                     WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT
                     EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"
     gene            5350. .6231
                     /gene="Z0989"
     CDS             5350. .6231
                     /gene="Z0989"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55142.1"
                     /db_xref="GI:12513757"
                     /translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV
                     KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGEISRKIEVMLNASPVSAAM
                     ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                     FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDIIGIRMDKINGESLLNISSLPAQ
                     AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                     SYHGGKQDLISVVLSKI"
     gene            6462. .7160
                     /gene="Z0990"
     CDS             6462. .7160
```

```
                     /gene="Z0990"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55143.1"
                     /db_xref="GI:12513758"
                     /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLRF
                     QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                     IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                     SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                     VSLGFQQVRLHPLL"
     gene            complement(7668. .8144)
                     /gene="ybhB"
                     /note="Z0992"
     CDS             complement(7668. .8144)
                     /gene="ybhB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                     residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                     MG1655: B0773"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55144.1"
                     /db_xref="GI:12513759"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(8203. .9492)
                     /gene="bioA"
                     /note="Z0993"
     CDS             complement(8203. .9492)

  Query Match             93.6%;  Score 23.4;  DB 1;  Length 13501;
  Best Local Similarity   96.0%;  Pred. No. 2;
  Matches   24;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          ||||||||||||||||||||| |||
Db   9689 TCGTCAGGTGCAGGTCAGCACCTTG 9713


RESULT  15
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .1245
                     /gene="ECs0753"
     CDS             79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
     gene            1245. .2114
                     /gene="ECs0754"
     CDS             1245. .2114
                     /gene="ECs0754"
                     /note="similar to SUCD_ECOLI gi|1786949 percent identity
                     100 in 289 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase alpha subunit"
                     /protein_id="BAB34177.1"
                     /db_xref="GI:13360213"
                     /translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
                     GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
                     LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
```

```
                AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
                AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
                LADIGEALKTVLK"
     gene       2781. .3686
                /gene="ECs0755"
     CDS        2781. .3686
                /gene="ECs0755"
                /note="probable transcriptional regulator,similar to
                transcriptional regulators e.g. glycine cleavage system
                transcription activator (gcv operon activator) -
                Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
                identity 31 in 300 aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transcription regulator"
                /protein_id="BAB34178.1"
                /db_xref="GI:13360214"
                /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
                EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
                PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
                LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
                NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
                EKSDSPHLQQFREWLLAKSVVPQEM"
     gene       complement(3720. .4322)
                /gene="ECs0756"
     CDS        complement(3720. .4322)
                /gene="ECs0756"
                /note="probable cob(I)alamin adenosyltransferase,similar
                to cob(I)alamin adenosyltransferases (corrinoid
                adenosyltransferases) e.g. [Escherichia coli]
                gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative cob(I)alamin adenosyltransferase"
                /protein_id="BAB34179.1"
                /db_xref="GI:13360215"
                /translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
                KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
                QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
                IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"
     gene       complement(4332. .5984)
                /gene="ECs0757"
     CDS        complement(4332. .5984)
                /gene="ECs0757"
                /note="probable fumarate hydratase,similar to fumarate
                hydratases e.g. fumarate hydratase class I, aerobic
                (fumarase) - Escherichia coli
                gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative fumarate hydratase"
                /protein_id="BAB34180.1"
                /db_xref="GI:13360216"
                /translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
                ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
                NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
                TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
                GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
                EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
                EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
                DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
                QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSIGGAAALLAQEYVKSLRCLEYPELG
                MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"
     gene       complement(6092. .7372)
                /gene="ECs0758"
     CDS        complement(6092. .7372)
                /gene="ECs0758"
                /note="probable transport protein,similar to
                glutamate/aspartate transport proteins (proton glutamate
                symport proteins) e.g. [Bacillus stearothermophilus]
                gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
                aa,also similar to C4-dicarboxylate transpor"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transport protein"
                /protein_id="BAB34181.1"
                /db_xref="GI:13360217"
                /translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
                TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
                NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
                FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
                VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
                GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
                LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
                AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"
     gene       complement(7533. .7850)
                /gene="ECs0759"
     CDS        complement(7533. .7850)
                /gene="ECs0759"
                /note="unknown"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="hypothetical protein"
                /protein_id="BAB34182.1"
```

Query Match 93.6%; Score 23.4; DB 1; Length 297816;
Best Local Similarity 96.0%; Pred. No. 1.3;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          ||||||||||||||||||||| |||
Db  92421 TCGTCAGGTGCAGGTCAGCACCTTG 92445
```

```
RESULT  16
SMABIO
LOCUS       SMABIO       7215 bp    DNA             BCT       04-FEB-1999
DEFINITION  S. marcescens biotin operon, complete sequence.
ACCESSION   D17468
VERSION     D17468.1  GI:402530
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-aminopelargonic acid synthetase; DAPA aminotransferase;
            DTB synthetase; KAPA synthetase; biotin operon; biotin synthetase;
            dethiobiotin synthetase.
SOURCE      Serratia marcescens (strain:Sr41) DNA.
  ORGANISM  Serratia marcescens
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Serratia.
REFERENCE   1  (bases 1 to 7215)
  AUTHORS   Sakurai,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1993) to the DDBJ/EMBL/GenBank databases. Naoki
            Sakurai, Tanabe Seiyaku Co., Ltd., Res Lab of Applied Biochemistry;
            2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan
            (E-mail:GFC01101@niftyserve.or.jp, nsakurai@ddbj.nig.ac.jp,
            Tel:048-433-2545, Fax:048-433-2540)
REFERENCE   2  (bases 1 to 7215)
  AUTHORS   Sakurai,N., Imai,Y., Akatsuka,H., Kawai,E., Komatsubara,S. and
            Tosa,T.
  TITLE     Complete nucleotide sequence of biotin operon of Serratia
            marcescens
  JOURNAL   Unpublished (1993)
COMMENT     Submitted (25-Aug-1993) to DDBJ by:
            Naoki Sakurai
            Res. Lab. of Applied Biochemistry
            Tanabe Seiyaku Co., Ltd.
            2-50 Kawagishi-2-chome
            Toda, Saitama 335
            Japan
```

Phone: 048-433-2545
Email: nsakurai@ddbj.nig.ac.jp
Fax: 048-433-2540.

```
FEATURES             Location/Qualifiers
     source          1. .7215
                     /organism="Serratia marcescens"
                     /strain="Sr41"
                     /db_xref="taxon:615"
     gene            complement(1353. .2630)
                     /gene="bioA"
     CDS             complement(1353. .2630)
                     /gene="bioA"
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase
                     (DAPA aminotransferase)"
                     /protein_id="BAA04284.1"
                     /db_xref="GI:402531"
                     /translation="MSVTASDLAFDQRHIWHPYTSMSRPLPCYPIESASGVELQLADG
                     RSLVDGMSSWWAAIHGYNHPRLNQAASQQLEKMSHVMFGGITHPAAISLCRRLVEMTP
                     EALQCVFLADSGSVAVEVSLKMALQYWQARGERRQRILTLRHGYHGDTFGAMSVCDPD
                     NSMHSLYQGYLAPHLFATAPQCRFDEEWREEDIAPFAALLEQHAGEVAAVILEPVVQG
                     AGGMRIYHPTYLKRVRDVVRVAIKLLLIADEIATGFGRTGKLFACEHAQVVPDILCLG
                     KALTGGYMTLSATLTTRHVAETISNGAAGCFMHGPTFMGNPLACAVADASLALLAENR
                     WQAQVSAIETQLKRELLPLGRCRRWPHVRVLGAIGVVEMREPVDVAELQRGFVERGVW
                     IRPFGKLIYLMPPYIIEAEQLSRLTAAVADAAR"
     gene            2716. .3756
                     /gene="bioB"
     CDS             2716. .3756
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="BAA04285.1"
                     /db_xref="GI:402532"
                     /translation="MMADRIHWTVGQAQALFDKPLLELLFEAQTVHRQHFDPRQVQVS
                     TLLSIKTGACPEDCKYCPQSPRYKTGLESERLMQVEQVLESARKAKANGSTRFCMGAA
                     WKNPHERDMPYLQQMVQGVKAMGMETCMTLGTLDGTQAERLAEAGLDYYNHNLDTSPE
                     FYGSIITTRSYQERLDTLDKVRDAGIKVCSGGIVGLGETVRDRAGLLVQLANLPKPPE
                     SVPINMLVKVKGTPLADNDDVDPFDFIRTIAVARIMMPSSYVRLSAGREQMNEQTQAM
                     CFMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTATEHGDNQQQQVLAKQLL
                     NADTAEFYNAAP"
     gene            3756. .4904
                     /gene="bioF"
     CDS             3756. .4904
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-aminopelargonic acid synthetase (KAPA
                     synthetase)"
                     /protein_id="BAA04286.1"
                     /db_xref="GI:402533"
                     /translation="MSWQQRIEQALAERRLNAAYRRRQTTEGGNGRQIRLGDRLYLNF
                     SGNDYLGLSQDARVIAAWQQGARYGVGNGGSGHVTGFSAAHQALEEQLAAWLGYPRAL
                     LFISYAANQAVLAALMQKGDRILADRLSHASLLEAAGAVAGRRAPVPAQSTAGLARIC
                     LAKPCDGQRLAVTEGLFSMDGDGAPLAELHRLTRAAGAWLMVDDAHGIGVRGEQGRGS
                     CWQQGVRPELLVATFGKAFGVSGAAVLCDEATAEYLLQFARHLIYSTAMPPAQACALQ
                     AALARIREGDDLRARLQDNIRRFRQGAAPLALTLTDSDTAIQPLLVGDNQRALDLATR
                     LRECGLWVSAIRPPTVPPGGARLRLLLTAAHQSQDIDRLLEVLNDVSQ"
     gene            4888. .5655
                     /gene="bioC"
     CDS             4888. .5655
                     /gene="bioC"
                     /function="synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="the product of bioC"
                     /protein_id="BAA04287.1"
                     /db_xref="GI:402534"
                     /translation="MTSANDTVNKQAVASAFSRAAGSYDAAAALQRDVGERLLGMGSS
                     HPGEQLLDAGCGTGYFSRMWRERGKRVTALDLAPGMLDVARQRQAAHHYLLGDIEQVP
                     LPDAAMDICFSSLVVQWCSDLPAALAELYRVTRPGGVILFSTLAAGSLQELGDAWQQV
                     DGERHVNAFLPLTQIRTACAAYRHELVTELRTLNYPDVMTLMRSLKGIGATHLHQGRE
                     GGLMSRGRLAALQAAYPCRQGQFPLSYHLAYGVIYRE"
     gene            5648. .6331
                     /gene="bioD"
     CDS             5648. .6331
                     /gene="bioD"
                     /function="conversion from 7,8-diaminopelargonic acid into
                     dethiobiotin"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase(DTB synthetase)"
                     /protein_id="BAA04288.1"
                     /db_xref="GI:912453"
                     /translation="MSKRWFVTGTDTEVGKSVASSALLQAANRAGYPSAGYKPVASGS
                     EMTAEGLRNGDALALQANSGVALDYDEVNPYVFAEPTSPHIVSADEGRPIDAARLSDG
                     LRRLEQRADWVLVEGAGGWFTPLSAEYTFADWVRQEQLPVILVVGIKLGCINHAVLTA
                     QAVQQAGLTLAGWIANDVGGAPGRRHQEYLATLRRMLPRRCWAKSRTCRRPERAPLGQ
                     YLDISLLAQ"
BASE COUNT     1347 a    2225 c    2367 g    1276 t
ORIGIN


  Query Match             89.6%;  Score 22.4;  DB 1;  Length 7215;
  Best Local Similarity   95.8%;  Pred. No. 6.4;
  Matches   23;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||| ||
Db   2830 CGTCAGGTGCAGGTCAGCACGCTG 2853


RESULT  17
AF250768
LOCUS       AF250768     4000 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosFS1 BioA-like protein (bioA) gene, partial
            cds; BioB-like protein (bioB) and KAPA-synthetase (bioF) genes,
            complete cds; and biotin biosynthesis protein BioC (bioC) gene,
            partial cds.
ACCESSION   AF250768
VERSION     AF250768.1  GI:12620095
KEYWORDS    .
SOURCE      uncultured bacterium pCosFS1.
  ORGANISM  uncultured bacterium pCosFS1
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 4000)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 4000)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .4000
                     /organism="uncultured bacterium pCosFS1"
                     /db_xref="taxon:143795"
                     /clone="pCosFS1"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     gene            complement(<1. .1177)
                     /gene="bioA"
     CDS             complement(<1. .1177)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="BioA-like protein"
                     /protein_id="AAG60558.1"
                     /db_xref="GI:12620098"
                     /translation="MSVTDSDLEFDQRHIWHPYTSMSHPLPCYPVEAASGVELQLADG
```

```
                     RRLIDGMSSWWAAIHGYNHPQLNLAATQQLEKMSHVMFGGITHPAANELCRKVVAMTP
                     EALQCVFLADSGSVAVEVSLKMALQYWQARGERRQRILTLRHGYHGDTFGAMSVCDPD
                     NSMHSLYQGYLAPHLFATAPQCGFDQAWDENDIAPFAALLEQHAGEVAAVILEPVVQG
                     AGGMRIYHPTYLERVRELCDRYQVLLIADEIATGFGRTGKLFACEHAEVVPDILCLGK
                     ALTGGYMTLSATLTTRHVAETISNGAAGCFMHGPTFMGNPLACAVANASLSLLAENRW
                     QQQVAAIETQLKQQLLPLKMQPKVADVRVLGAIGVVEMREPVDMAQLQRGFVERGVWI
                     "
     gene            1265. .2305
                     /gene="bioB"
     CDS             1265. .2305
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="BioB-like protein"
                     /protein_id="AAG60559.1"
                     /db_xref="GI:12620099"
                     /translation="MMADRIHWTVGLAQTLFDKPLLELLFEAQTVHRQHFDPRQVQVS
                     TLLSIKTGACPEDCKYCPQSSRYKTGLESERLMQVEQVLESARKAKANGSTRFCMGAA
                     WKNPHERDMPYLQQMVQGVKAMGMETCMTLGTLDGSQAERLAEAGLDYYNHNLDTSPE
                     FYGNIITTRSYQERLDTLGKVRGAGIKVCSGGIVGLGETVXXRAGLLVQLANLPTPPE
                     SVPINMLVKVKGTPLADNDDVDPFDFIRTIAVARIMMPSSYVRLSAGREQMNEQTQAM
                     CFMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTETEHGDNQQQEVLAARLM
                     QADTAEFYNAAL"
     gene            2305. .3456
                     /gene="bioF"
     CDS             2305. .3456
                     /gene="bioF"
                     /note="7-keto-8-aminopelargonic acid synthetase; similar
                     to BioF"
                     /codon_start=1
                     /transl_table=11
                     /product="KAPA-synthetase"
                     /protein_id="AAG60556.1"
                     /db_xref="GI:12620096"
                     /translation="MSWQQRIEQALEERXQSAAYRHRLVIKGGNGRYLQLNDQRYLNF
                     SGNDYLGLSQNAQVIAAWQQGAQQYGVGSGGSGHVTGFCTAHQALEQQLAQWLGYPRS
                     LLFISGYAANQALLAALMQGEDRILADRLSHASLLEAAAHSPAQLRRFRHNQPQALAD
                     LLAKPCSGQTLAVTEGVFSMDGDSAPLAELYRLTRDAGAWLMVDDAHGIGVQGEQGRG
                     SCWQQGVRPELLVVTFGKAFGLSGAAVLCDESTAEYLLQFARHLIYSTAMPPAQACAL
                     QAALHCIQQGDELRTRLNTNIQRFRQGAAQLSLTLTESLTAIQPLLVGDNQRALDLAQ
                     HLRGQGLWVSAIRPPTVPPGGARLRITLTAAHQPQDIDRLLEVLHDVSV"
     gene            3440. .>4000
                     /gene="bioC"
     CDS             3440. .>4000
                     /gene="bioC"
                     /note="similar to BioC; probably reaction step prior to
                     synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin biosynthesis protein BioC"
                     /protein_id="AAG60557.1"
                     /db_xref="GI:12620097"
                     /translation="MTSAFEQVNKQAVASAFSRAAVGYDAAAVLQRDVGERLLGMGQE
                     HQGEQLLDAGCGTGYFSRRWRELGKQVTALDLAPGMLAFARQQQAADHYLLGDIENVP
                     LGDAAVDISFSSLVVQWCSDLPRALAELYRVTRPGGVILFSTLAEGSLHELGDAWQQV
                     DGERHVNDFLPLAQIEAACAGYRHHLQ"
BASE COUNT      794 a    1164 c    1236 g     802 t      4 others
ORIGIN


  Query Match             83.2%;  Score 20.8;  DB 1;  Length 4000;
  Best Local Similarity   91.7%;  Pred. No. 37;
  Matches    22;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           |||||||||||||| |||||| ||
Db    1379 CGTCAGGTGCAGGTGAGCACGCTG 1402


RESULT  18
EHBIOTOP2
LOCUS       EHBIOTOP2    1041 bp    DNA             BCT       04-NOV-1996
DEFINITION  Erwinia herbicola biotin synthetase (bioB) gene, complete cds.
ACCESSION   U38648
VERSION     U38648.1  GI:1228112
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Erwinia herbicola strain=Eho10.
  ORGANISM  Pantoea agglomerans
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Pantoea.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Wu,C.H., Chen,H.Y. and Shiuan,D.
  TITLE     Isolation and characterization of the Erwinia herbicola bio operon
            and the sequences of the bioA and bioB genes
  JOURNAL   Gene 174 (2), 251-258 (1996)
  MEDLINE   97045821
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Shiuan,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-1995) David Shiuan. Biology, Natl. Sun Yat-Sen
            University, Kaohsiung, Taiwan, ROC
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Pantoea agglomerans"
                     /strain="Eho10"
                     /db_xref="taxon:549"
     gene            1. .1041
                     /gene="bioB"
     CDS             1. .1041
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAC44534.1"
                     /db_xref="GI:1228114"
                     /translation="MAHHARWTLSQVTELFDKPLLDLLFEAQTIHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLETERLMEVEQVLDSARKAKNAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLDETQAQRLASAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLDKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSFVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDVQLFRKLGLNPQQTDVMTGDNEQQQKLEEQIFN
                     ADTDQFYNAAAL"
BASE COUNT      230 a    322 c    308 g    181 t
ORIGIN


  Query Match             80.8%;  Score 20.2;  DB 1;  Length 1041;
  Best Local Similarity   88.0%;  Pred. No. 84;
  Matches    22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           ||| ||||| |||||||||||| |||
Db     111 TCGCCAGGTTCAGGTCAGCACCTTG 135


RESULT  19
AE001862/c
LOCUS       AE001862   213732 bp    DNA             BCT       22-NOV-1999
DEFINITION  Deinococcus radiodurans R1 section 1 of 2 of the complete
            chromosome 2.
ACCESSION   AE001862 AE001825
VERSION     AE001862.1  GI:6460468
KEYWORDS    .
SOURCE      Deinococcus radiodurans.
  ORGANISM  Deinococcus radiodurans
            Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
REFERENCE   1  (bases 1 to 213732)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et,al.
  TITLE     Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
  JOURNAL   Science 286 (5444), 1571-1577 (1999)
```

```
  MEDLINE   20036896
REFERENCE   2  (bases 1 to 213732)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .213732
                     /organism="Deinococcus radiodurans"
                     /strain="R1"
                     /db_xref="taxon:1299"
                     /chromosome="2"
     gene            653. .1435
                     /gene="DRA0001"
     CDS             653. .1435
                     /gene="DRA0001"
                     /note="similar to GB:D26185 SP:P37522 PID:467381
                     PID:580906 GB:AL009126 percent identity: 54.96; identified
                     by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="chromosome partitioning ATPase, putative, ParA
                     family"
                     /protein_id="AAF12301.1"
                     /db_xref="GI:6460595"
                     /translation="MVSAVKTLTVFNHAGGAGKTSLTLNVGYELARGGLRVLLLDLDP
                     QANLTGWLGISGVTREMTVYPVAVDGQPLPSPVKAFGLDVIPAHVSLAVAEGQMMGRV
                     GAQGRLRRALAEVSGDYDVALIDSPPSLGQLAILAALAADQMIVPVPTRQKGLDALPG
                     LQGALTEYREVRPDLTVALYVPTFYDARRRHDQEVLADLKAHLSPLARPVPQREAVWL
                     DSTAQGAPVSEYAPGTPVHADVQRLTADIAAAIGVAYPGENA"
     gene            1432. .2313
                     /gene="DRA0002"
     CDS             1432. .2313
                     /gene="DRA0002"
                     /note="similar to GB:L34077 GB:X85964 PID:407375
                     PID:757761 PID:1163135 percent identity: 57.34; identified
                     by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="chromosome partitioning protein, ParB family"
                     /protein_id="AAF12300.1"
                     /db_xref="GI:6460594"
                     /translation="MTRRRPERRRDLLGLLGETPVDLSQANDIRALPVNELKVGSTQP
                     RRSFDLERLSELAESIRAHGVLQPLLVRSVDGQYEIVAGERRWRAAQLAGLAEVPVVV
                     RQLSNEQARAAALIENLQRDNLNVIDEVDGKLELIALTLGLEREEARKRLMQLLRAVP
                     GDEHEQLDQVFRSMGETWRTFAKNKLRILNWPQPVLEALRAGLPLTLGSVVASAPPER
                     QAELLKLAQNGASRSQLLQALQTPSQTSAVTPEHFAKVLSSKRFLSGLDTPTREALDR
                     WLARMPERVRQAIDEQS"
     gene            complement(2418. .3881)
                     /gene="DRA0003"
     CDS             complement(2418. .3881)
                     /gene="DRA0003"
                     /note="similar to GB:M88334 SP:P25526 PID:147901 GB:U00096
                     PID:1789015 percent identity: 69.60; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="succinate-semialdehyde dehydrogenase"
                     /protein_id="AAF12294.1"
                     /db_xref="GI:6460588"
                     /translation="MTTTDLRTTYSSVTRSQAYFDGEWRNAPRNFEVRHPGNGEVIGE
                     VADCTPTDARQAIDAAEVALREWRQVNPYERGKILRRWHDLMFEHKEELAQLMTLEMG
                     KPISETRGEVHYAASFIEWCAEEAGRIAGERINLRFPHKRGLTISEPVGIVYAVTPWN
                     FPAGMITRKAAPALAAGCVMILKPAELSPMTALYLTELWLKAGGPANTFQVLPTNDAS
                     ALTQPFMNDSRVRKLTFTGSTEVGRLLYQQAAGTIKRVSLELGGHAPFLVFDDADLER
                     AASEVVASKFRNSGQTCVCTNRVYVQRGVAEEFIRLLTEKTAALQLGDPFDEATQVGP
                     VVEQAGLDKVQRQVQDALTKGAQATTGGQVSSGLFFQPTVLVDVAPDSLILREETFGP
                     VAPVTIFDTEEEGLRLANDSEYGLAAYAYTRDLGRAFRIAEGLEYGIVGINDGLPSSA
                     APHVPFGGMKNSGVGREGGHWGLEEYLETKFVSLGLS"
     gene            4024. .4668
                     /gene="DRA0004"
     CDS             4024. .4668
                     /gene="DRA0004"
                     /note="similar to GP:2924378 percent identity: 55.29;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF12299.1"
                     /db_xref="GI:6460593"
                     /translation="MNSLSSLAAVLLMAGRSTRMGQPKALLPLGGMPLCRHAAQTLVE
                     AGYGEVWAVLPPGETGDAVRAALQGLPLRWVTNPQPAQGLLSSFQAAGEALWAESAAP
                     AVVFTLADMPLVRRDTHQALGTVFGAEQPAAIITRYGEVAAPPTLLRRDLLPRLHEFP
                     PSDHGPRTLLRELGRELKKIDRPAAELLDIDTPEAWAQAQQLFAEVTPGAQDGP"
     gene            complement(4719. .5897)
                     /gene="DRA0005"
     CDS             complement(4719. .5897)
                     /gene="DRA0005"
                     /note="similar to GB:L33464 SP:P47734 PID:496118 percent
                     identity: 74.23; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="alcohol dehydrogenase, zinc-containing"
                     /protein_id="AAF12293.1"
                     /db_xref="GI:6460587"
                     /translation="MKAIVWQGVNKVGVETVPDPTLLLPTDAIVKITSTAICGSDLHL
                     LDGYIPSMETGDILGHEFMGEVVEVGSEVRKLKVGDRVVVPFNIACGVCDPCKRGFFS
                     ACDNSNPNHRMAEALYGATSGGALFGYSHMYGGYAGGQAQYVRVPFADVGPFKIESGL
                     KDEQVLFLTDIFPTGYQAAEQCNIVPGRDVVAVFGAGPVGQFAARSAQMLGAAHVIVI
                     DRVPERLAMAEAAGCQTINYEKEDVLLSLREATGGRGPDHVIDAVGMEAHGHGPGSTM
                     DKVQQNLKLTFDRITALRWAILSCAKGGTVSMPGVYGGLVDKMPIGAAFAKGLIFRMG
                     QTHTHRHIPSLLGRIEAGQIDPSFVITHRASLDQAPELYKTFRDKQDCCIKVVLNPWA
                     "
     gene            complement(5894. .6700)
                     /gene="DRA0006"
     CDS             complement(5894. .6700)
                     /gene="DRA0006"
                     /note="similar to PID:1653488 percent identity: 48.46;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF12292.1"
                     /db_xref="GI:6460586"
                     /translation="MDRRSSPQLRTMTNWNREQPAPHAGATGPNSSPGATMQPLERSL
                     TGAVGAALAAAGLRSRSGLEKVVLTGLGAGLIAVAARGQNPLATALKIEQNDEGETVV
                     RDAVTIGKPAEQLYAVWRDLPGLPLLMTHLRSVEVLDDKRSRWTVEAPAPLGTVSWEA
                     ELTADEPGKRIAWRSLPGARIENSGEVLFRPAPGARGTEVVVRLTYRPPGGSAGAVIA
                     RMFNQEPSQQLRDDLMRFKREQELGHAPTAQGQSSGRAALTPPNKRGDLV"
     gene            6839. .7744
                     /gene="DRA0007"
     CDS             6839. .7744
                     /gene="DRA0007"
                     /note="similar to GB:AL009126 percent identity: 62.98;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter, ATP-binding protein, putative"
                     /protein_id="AAF12291.1"
                     /db_xref="GI:6460585"
                     /translation="MITTFEQVSKTYGHVTALSDFNLTLRTGELTALLGPNGAGKSTA
                     IGLLLGLSAPSAGQVRVLGADPRRNDVRARIGAMPQESALPAGLTVREAVTLFASFYP
                     APLGVDEALALADLGPVAGRRAAQLSGGQKRRLAFALAVVGDPELLLIDEPTTGMDAQ
                     SRAAFWEAVTGLRARGRTILLTTHYLEEAERTADRVVVMNGGRILADDTPQGLRSGVG
                     GARVSFVSDLVQAELERLPGVSAVQVDAAGRADLRTSVPEALLAALIGSGTTFSDLEV
                     RRATLEEAYLQLTGPQDMTAVTRSA"
     gene            7861. .8652
                     /gene="DRA0008"
     CDS             7861. .8652
```

```
                     /gene="DRA0008"
                     /note="similar to GB:AL009126 percent identity: 52.46;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF12290.1"
                     /db_xref="GI:6460584"
                     /translation="MTTLPNVPARRTSALPVLAQLTLGELRRLLRSPMFTVGAIGFPV
                     MFFALFGLPAVQEYGATDPHVGPVILTQFAAYSLLSLALFSFGAAVATERSGGWLRLL
                     RSSPLPVPLYFVSKTLAALAFGAVSLALLYAFAHFAGGVTLPLGLALLLAGKLLLGMI
                     PLVALGLCIGFLASPQAAQILANILSVVMSFASGLFVPLDQLPGFVQQIAPLLPAYHV
                     SQIATNTVSGQTASEPAHWLALAAFTLVFGTLAVWGLKRDESREG"
     gene            8809. .9819
                     /gene="DRA0009"
     CDS             8809. .9819
                     /gene="DRA0009"
                     /note="similar to GB:AL009126 percent identity: 55.80;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF12289.1"
                     /db_xref="GI:6460583"
                     /translation="MLFWGTLVAFVAVYVFTFGKRDPDDRWAYVGWLSALLMWALGDW
                     LYGPSALSFLVYGGSLIGWQRRQWVALGGAAINALLMVGRLLTLPERSQGDFIFLLFI
                     LASAYANHATYRGLRARRRLAQVQLEKEKLAADAERERIARDLHDLLGHTLSVIVLKS
                     ELAGKLAEKHPQRAAAEIREVERISREALSEVRAAVQGYRGSGLAAELARAKVALDAA
                     GIRLVITDTLPELPADLESTLAMVLREAVTNVVRHSGASELRLTLRGQANGFQLTLQD
                     DGRGGDAPEGTGLNSMRERLRAMGGTFERDGHTGTRLQAFVPLSAGTAGLKAPELLEK
                     TS"
     gene            9816. .10451
                     /gene="DRA0010"
     CDS             9816. .10451
                     /gene="DRA0010"
                     /note="similar to GB:AL009126 percent identity: 67.84;

  Query Match              77.6%;  Score 19.4;  DB 1;  Length 213732;
  Best Local Similarity    95.2%;  Pred. No. 93;
  Matches   20;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtcagcacgtt 24
           |||||||||||||||||||| |
Db   52231 TCAGGTGCAGGTCAGCACGGT 52211


RESULT  20
AC019891/c
LOCUS       AC019891    50195 bp    DNA             HTG       03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC019891
VERSION     AC019891.1  GI:6665006
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 50195)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211291 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .50195
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    14230 a  11038 c  10533 g  14394 t
ORIGIN


  Query Match              74.4%;  Score 18.6;  DB 2;  Length 50195;
  Best Local Similarity    84.0%;  Pred. No. 2.6e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           ||| |||||||||||| ||| | ||
Db   37133 TCGACAGGTGCAGGTCTGCAAGGTG 37109


RESULT  21
AC003055
LOCUS       AC003055    73076 bp    DNA             INV       04-NOV-1997
DEFINITION  Drosophila melanogaster (P1 DS06332 (D91)) DNA sequence, complete
            sequence.
ACCESSION   AC003055 AC001994 AC002947 AC002944 AC002943 AC002952 AC001992
            AC002945 AC002946 AC001988 AC001995 AC002951 AC001991 AC002942
            AC001993 AC002948 AC002953 AC001990 AC001989 AC002949 AC002950
VERSION     AC003055.1  GI:2584828
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (Subclones in pOT2 from P1 clone DS06332
            (D91)) DNA.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 73076)
  AUTHORS   Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Doyle,C.M.,
            Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Kim,S.H., Ko,C.L., Li,M., Lomotan,M.A., Mazda,P.,
            Mok,M.S., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D.,
            Santos,R.F., Snir,E., Stevko,V., Subramanian,S., Towne,B.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and
            Kimmel,B.
  TITLE     Sequencing of Drosophila chromosome 2L, region 22F
  JOURNAL   Unpublished (1997)
REFERENCE   2  (bases 1 to 73076)
  AUTHORS   Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Doyle,C.M.,
            Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Kim,S.H., Ko,C.L., Li,M., Lomotan,M.A., Mazda,P.,
            Mok,M.S., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D.,
            Santos,R.F., Snir,E., Stevko,V., Subramanian,S., Towne,B.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and
            Kimmel,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1997) Berkeley Drosophila Genome Project, MS
            74-157, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 74-157
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to  other sequences, please visit our sequence
            archive Web site (http://www-hgc.lbl.gov/sequence-archive.html) or
            send email to drosophila@genome.lbl.gov.
            Library location: 66-92.
            This P1 was assembled from the following subclones: 2_c1
            (AC001994), 2_b12 (AC002947), 1_g6 (AC002944), 1_d4 (AC002943),
            2_b5, 2_e4 (AC002952), 1_b4, 2_a1, 2_a2, 1_h6 (AC001992), 1_h8
            (AC002945), 2_a8 (AC002946), 1_b2 (AC001988), 2_c5 (AC001995), 2_d5
            (AC002951), 1_h3 (AC001991), 1_d3 (AC002942), 2_a9 (AC001993),
            2_c10 (AC002948), 2_g9 (AC002953), 1_g7 (AC001990), 1_b5
            (AC001989), 2_c3 (AC002949), 2_h2, 2_c4 (AC002950), 2_b6.
FEATURES             Location/Qualifiers
     source          1. .73076
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
```

```
                     /map="22F2-22F4"
                     /clone="P1 DS06332 (D91)"
BASE COUNT    21203 a  14890 c  15769 g  21214 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 3;  Length 73076;
  Best Local Similarity   84.0%;  Pred. No. 2.5e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          ||| |||||||||||| ||| | ||
Db  17969 TCGACAGGTGCAGGTCTGCAAGGTG 17993


RESULT  22
AC092189/c
LOCUS       AC092189   168540 bp    DNA             INV       26-JUN-2001
DEFINITION  Drosophila melanogaster, chromosome 2L, region 22E-22F, BAC clone
            BACR34M11, complete sequence.
ACCESSION   AC092189
VERSION     AC092189.1  GI:14550291
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 168540)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Sequencing of Drosophila chromosome 2L, region 22E-22F
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168540)
  AUTHORS   Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
            Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
            Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
            Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
            Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
            Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
            Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
            McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
            Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
            Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
            Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
            Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2001) Berkeley Drosophila Genome Project, MS
            64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            This sequence was assembled using end sequences from a whole genome
            shotgun and from subclones of this BAC and its neighboring clones.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .168540
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="22E-22F"
                     /clone="BACR34M11 (D1362)"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6)"
BASE COUNT    47515 a  37192 c  36096 g  47737 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 3;  Length 168540;
  Best Local Similarity   84.0%;  Pred. No. 2.2e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          ||| |||||||||||| ||| | ||
Db 137787 TCGACAGGTGCAGGTCTGCAAGGTG 137763


RESULT  23
AC026904
LOCUS       AC026904   175369 bp    DNA             HTG       30-JUL-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-567J20 map 6, *** SEQUENCING
            IN PROGRESS ***, 1 ordered pieces.
ACCESSION   AC026904
VERSION     AC026904.5  GI:14547859
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175369)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 6, clone RP11-567J20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175369)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 25, 2001 this sequence version replaced gi:14333958.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
```

```
           -------------- Project Information
               Center project name: L7093
               Center clone name: 567_J_20
           --------------.
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 1 contigs. Gaps between the contigs
           * are represented as runs of N. The order of the pieces
           * is believed to be correct as given, however the sizes
           * of the gaps between them are based on estimates that have
           * provided by the submittor.
           * This sequence will be replaced
           * by the finished sequence as soon as it is available and
           * the accession number will be preserved.
           *        1   175369: contig of 175369 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175369
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6"
                     /clone="RP11-567J20"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    51321 a  39056 c  37390 g  47602 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 2;  Length 175369;
  Best Local Similarity   84.0%;  Pred. No. 2.2e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           ||||| ||||||| ||||| ||||
Db 137375 TCGTCTTGTGCAGGACAGCAGGTTG 137399


RESULT  24
AE003582/c
LOCUS       AE003582   295312 bp    DNA             INV       04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386046 section 9
            of 16, complete sequence.
ACCESSION   AE003582 AE002638
VERSION     AE003582.2  GI:10727369
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 295312)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 295312)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7295900.
FEATURES             Location/Qualifiers
     source          1. .295312
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
     mRNA            complement(join(16713. .17645,17816. .>18127))
                     /gene="CG3515"
                     /product="CT11855"
                     /db_xref="FLYBASE:FBan0003515"
                     /db_xref="FLYBASE:FBgn0031431"
     gene            complement(<16713. .>18127)
                     /gene="CG3515"
                     /map="22F3-23A1"
                     /db_xref="FLYBASE:FBan0003515"
                     /db_xref="FLYBASE:FBgn0031431"
     CDS             complement(join(16713. .17645,17816. .18043))
                     /gene="CG3515"
                     /note="CG3515 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0003515"
                     /db_xref="FLYBASE:FBgn0031431"
                     /protein_id="AAF51244.1"
                     /db_xref="GI:7295945"
                     /translation="MRQFRVAKVFPETKHHTLAMEFSSDGLRLLVCDHSALSIFSSTR
                     QVQVCKVPMHHYLPDVACFVQQDTRVLHSTSKHDYTIRCLDLNTRHCVRVFGGHAKTI
                     NRLASQPGGENVFISAGRDDQVYMWDIRVKTHTHRIKNLRRPLCAFDPEGLFLATSSG
                     TEGIQIRDVRMMGGEPSRKFVYQVNAKANWTQLQFAPNGKSLLLSTDHSWCFSVSAID
                     GTYQQSFTGYSNKLRLPLDATYTPDSQFILSGADKGRIHIWRAADGFPVAVLKGNNMG
                     PVRCMRFNPRATMFVSSDSRIVFWMPMANGVYDWVKRLKPGETISFTDETVQEEAVKA
                     PRTIPMPKPKIETVWKIKRLNRKRACTGDEIVDLTNALHNNDSMEDGEIAED"
     mRNA            complement(join(28221. .28381,28449. .28712,28820. .28943,
                     29016. .29435,29487. .29838,29893. .30164,30434. .30564,
                     30620. .>30841))
                     /gene="Cyp309a1"
                     /product="CT28097"
                     /db_xref="FLYBASE:FBan0009964"
                     /db_xref="FLYBASE:FBgn0031432"
     gene            complement(<28221. .>30841)
                     /gene="Cyp309a1"
                     /note="CG9964"
                     /map="22F4-23A1"
                     /db_xref="FLYBASE:FBan0009964"
                     /db_xref="FLYBASE:FBgn0031432"
     CDS             complement(join(28221. .28381,28449. .28712,28820. .28943,
                     29016. .29435,29487. .29838,29893. .30092))
                     /gene="Cyp309a1"
                     /note="Cyp309a1 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0009964"
                     /db_xref="FLYBASE:FBgn0031432"
```

```
                /protein_id="AAF51243.2"
                /db_xref="GI:10727370"
                /translation="MFTLVGLCLTIVHVAFAVVYFYLTWYHKYWDKRGVVTAEPLTIL
                GSYPGILINKSRSLILDVQDVYNKYKDKYRTVGTFITRQPQLLVLDPALAHEILVDKF
                SHFRDTITSSFVGHNPDDKYVAGSPFFSAGDKWKRLRSENVGGLTPSRLKMAYSIWEQ
                SGRKLVEYIERARREQGDIIETRDLAYRFTANAMADFIWGIDAGSLSGKVGEIGDFQK
                TSTDWSAHAFSSMIRFNKTLVAIFVRKLFSMRFFTKATDEFFLRLTQDAVNLRQGGSG
                EGRTDYLSHLIQLQQRGNSIHDSVGHALTVHLDGFETSGAVLYHMLYSLSEHHEEQEK
                LRSEILEALASEGQISYDQINNLPYLDQCFNESLRLTTPIGFFMRICTKPTQINLGDD
                KTLDLEPGVTVMVPAYQYHHDNDIYPEASEFRPDRFENGAASVLTKRGCFLPFGDGPR
                ICLGMRVGQLSVKTAIVHILSNYQVEQMKKVPLGADSGMGIFLNGDVELKYTKLQK"
mRNA            complement(join(<37517. .37711,37769. .38191,38253. .38604,
                38658. .>38947))
                /gene="CG18559"
                /product="CT41535"
                /db_xref="FLYBASE:FBan0018559"
                /db_xref="FLYBASE:FBgn0031433"
                /evidence=not_experimental
gene            complement(<37517. .>38947)
                /gene="CG18559"
                /map="23A1-23A1"
                /db_xref="FLYBASE:FBan0018559"
                /db_xref="FLYBASE:FBgn0031433"
                /evidence=not_experimental
CDS             complement(join(37517. .37711,37769. .38191,38253. .38604,
                38658. .38947))
                /gene="CG18559"
                /note="CG18559 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0018559"
                /db_xref="FLYBASE:FBgn0031433"
                /evidence=not_experimental
                /protein_id="AAF51242.1"
                /db_xref="GI:7295943"
                /translation="MMPERFSFDFRPIDQYWTRAKGACSNTKRGNMYILASLALILLH
                LLVLPIYLYLTWHHKYWRKRGLVTARPLTLLGTYPGLLTRKSNLVFDVQKIYDKYKGK
                HRAVGVFVTRQPQILVLDPELAHEVLVSNFRCYKDSLQSSYLRHAKWDKYARLNPFWA
                SGQSWRRLRTDAQAGISGSRLRQAYNIWEQGGQMLTEYMTQQVAEKNNILETRDLCFR
                YTAHVMADFIWGIDAGTLTRPMEQPNKVQEMASKWTSYAFYMLTLFMATIVAPCSRLL
                LRFRFYPKETDEFFSNLTKESIELRLKAGDSTRTDYLSHLLQLRDQKQATHDDLVGHA
                LTVMLDGYDTSGTALLHALYYLAENPAVQQKLRVEILSCMASEKSLDFEKLSSLQYLE
                QVIYGKHIYRLCQFHHIIYIVKLLFYL"
mRNA            complement(join(39362. .40140,40214. .40753,40814. .40989))
                /gene="Rad1"
                /product="CT10890"
                /db_xref="FLYBASE:FBan0003240"
                /db_xref="FLYBASE:FBgn0026778"
gene            complement(<39362. .>40989)
                /gene="Rad1"
                /note="CG3240"
                /map="23A1-23A1"
                /db_xref="FLYBASE:FBan0003240"
                /db_xref="FLYBASE:FBgn0026778"
CDS             complement(join(40320. .40753,40814. .40919))
                /gene="Rad1"
                /note="Rad1 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0003240"
                /db_xref="FLYBASE:FBgn0026778"
                /protein_id="AAF51241.1"
                /db_xref="GI:7295942"
                /translation="MASESENTTSYFVPHLVDQVFELYRSKAESNSLTKEERGQFFYD
                LSILLGQELVLRSLHLLDDHNFSFFHAKNNRSVCVVEISKGTEYFRLIPGVSFCKCEF
                FQCHVLQLPRGILYQDLPSGEPGILEDWSEESRVSYTCQHILALRLHQFLKHTGGKTT
                EKILKKEDIKELRTDVFRD"
mRNA            complement(join(41389. .42510,42575. .>42893))
                /gene="CG3227"
                /product="CT10839"
                /db_xref="FLYBASE:FBan0003227"
                /db_xref="FLYBASE:FBgn0031434"
gene            complement(<41389. .>42893)
                /gene="CG3227"
                /map="23A1-23A1"
                /db_xref="FLYBASE:FBan0003227"
                /db_xref="FLYBASE:FBgn0031434"
CDS             complement(join(41389. .42510,42575. .42583))
                /gene="CG3227"
                /note="CG3227 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0003227"
                /db_xref="FLYBASE:FBgn0031434"
                /protein_id="AAF51240.1"
                /db_xref="GI:7295941"
                /translation="MENQYQRIVSATAREDSQARRSKPLFLNAWSQTSSVDFEVLRSI
                SAPDPQVEVENRALRDKVRYLEAKLQQHKDLLSQIHATSARMQQASSLLAESRPPTPP
                AVQSHHILTPPSSEICAPAQNPQILDYKIISAPDDADAIEIRLAAESLNSLSTSADSD
                RLEICLGDENHQQSNHHNSQQQYRISNGIKRDGSSESADTPLQFIKRRKLQEIQLQEE
                IPRQNIKLPAKPQVKARNGSFTDLNKGQQQNMDNVMVSIGPNNTCVPASVFENINWSV
                CSLATRKLLVTIFDRETLATHSMTGKPSPAFKDQDKPLKRMLDPGKIQDIIFAVTHKC
                NASEKEVRNAITTKCADENKMMKIQNVKRRSSGIKHEKENII"
mRNA            43612. .45117

  Query Match             74.4%;  Score 18.6;  DB 3;  Length 295312;
  Best Local Similarity   84.0%;  Pred. No. 2e+02;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           ||| |||||||||||| ||| | ||
Db   17915 TCGACAGGTGCAGGTCTGCAAGGTG 17891


RESULT  25
HSDJ911I5
LOCUS       HSDJ911I5   72289 bp    DNA             PRI       23-FEB-2001
DEFINITION  Human DNA sequence from clone RP5-911I5 on chromosome 20q13.11-13.2
            Contains ESTs, STSs and GSSs, complete sequence.
ACCESSION   AL050316
VERSION     AL050316.15  GI:7406727
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 72289)
  AUTHORS   Howden,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 3, 2000 this sequence version replaced gi:7378749.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            IMPORTANT: This sequence is not the entire insert of clone
            RP5-911I5 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP5-911I5 is at 72289 in this sequence.
            The true right end of clone RP11-80K6 is at 100 in this sequence.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
```

```
            assembly was confirmed by restriction digest. RP5-911I5 is from the
            library RPCI-5 constructed by the group of Pieter de Jong. For
            further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .72289
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="q13.11-13.2"
                     /clone="RP5-911I5"
                     /clone_lib="RPCI-5"
     misc_feature    complement(1. .96)
                     /note="match: GSS: Em:AQ600021"
     repeat_region   120. .320
                     /note="MIR repeat: matches 66. .262 of consensus"
     repeat_region   328. .693
                     /note="THE1C repeat: matches 1. .371 of consensus"
     repeat_region   1311. .1534
                     /note="MIR repeat: matches 18. .262 of consensus"
     repeat_region   1734. .1889
                     /note="MER5A repeat: matches 31. .189 of consensus"
     repeat_region   4079. .4110
                     /note="16 copies 2 mer tt 87% conserved"
     repeat_region   4971. .5442
                     /note="L1PA7 repeat: matches 5675. .6145 of consensus"
     misc_feature    complement(6296. .6572)
                     /note="match: GSS: Em:B31666"
     repeat_region   6816. .6999
                     /note="AluJb repeat: matches 120. .303 of consensus"
     repeat_region   7323. .7456
                     /note="FLAM_C repeat: matches 2. .122 of consensus"
     repeat_region   7746. .7785
                     /note="5 copies 8 mer tacacata 85% conserved"
     repeat_region   8123. .8239
                     /note="AluJo/FRAM repeat: matches 181. .297 of consensus"
     repeat_region   9860. .10147
                     /note="AluSx repeat: matches 1. .288 of consensus"
     repeat_region   10520. .10705
                     /note="MIR repeat: matches 2. .191 of consensus"
     misc_feature    complement(11203. .11696)
                     /note="match: GSS: Em:AQ558975"
     repeat_region   11384. .11630
                     /note="L1ME repeat: matches 5579. .5819 of consensus"
     misc_feature    11697. .12071
                     /note="match: GSS: Em:AQ017203"
     repeat_region   12358. .12453
                     /note="12 copies 8 mer aggaagaa 79% conserved"
     repeat_region   12363. .12454
                     /note="23 copies 4 mer ggaa 81% conserved"
     misc_feature    complement(13112. .13596)
                     /note="match: GSS: Em:AQ154640"
     repeat_region   13252. .13483
                     /note="MIR repeat: matches 7. .243 of consensus"
     repeat_region   13883. .14071
                     /note="MIR repeat: matches 64. .261 of consensus"
     repeat_region   14118. .14425
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   14451. .14547
                     /note="L2 repeat: matches 2614. .2710 of consensus"
     repeat_region   14638. .14848
                     /note="MIR repeat: matches 22. .233 of consensus"
     repeat_region   14863. .15592
                     /note="L2 repeat: matches 1917. .2732 of consensus"
     repeat_region   16139. .16203
                     /note="L2 repeat: matches 2577. .2664 of consensus"
     repeat_region   16736. .16776
                     /note="L2 repeat: matches 2657. .2697 of consensus"
     repeat_region   16937. .16980
                     /note="11 copies 4 mer acat 77% conserved"
     repeat_region   18563. .18688
                     /note="AluSg/x repeat: matches 186. .310 of consensus"
     repeat_region   19350. .19611
                     /note="AluSp repeat: matches 41. .313 of consensus"
     repeat_region   19917. .20052
                     /note="L2 repeat: matches 2617. .2742 of consensus"
     repeat_region   20053. .20248
                     /note="AluSg/x repeat: matches 91. .285 of consensus"
     repeat_region   20249. .20291
                     /note="L2 repeat: matches 2578. .2617 of consensus"
     repeat_region   20292. .20582
                     /note="AluSc repeat: matches 1. .291 of consensus"
     repeat_region   20583. .20942
                     /note="L2 repeat: matches 2160. .2578 of consensus"
     repeat_region   22096. .22252
                     /note="MIR repeat: matches 24. .191 of consensus"
     repeat_region   22924. .23084
                     /note="MIR repeat: matches 32. .201 of consensus"
     repeat_region   23273. .23453
                     /note="L1ME repeat: matches 5603. .5802 of consensus"
     repeat_region   23747. .23956
                     /note="MER20 repeat: matches 6. .218 of consensus"
     repeat_region   24553. .24771
                     /note="MER46A repeat: matches 1. .236 of consensus"
     repeat_region   25172. .25464
                     /note="AluJo repeat: matches 1. .302 of consensus"
     misc_feature    complement(25399. .25912)
                     /note="match: GSS: Em:AQ381838"
     misc_feature    complement(25464. .25902)
                     /note="match: GSS: Em:AQ121579"
     repeat_region   25491. .25613
                     /note="MLT1I repeat: matches 287. .408 of consensus"
     repeat_region   25759. .25858
                     /note="MLT1I repeat: matches 1. .90 of consensus"
     repeat_region   26095. .26289
                     /note="MIR repeat: matches 3. .233 of consensus"
     repeat_region   27202. .27364
                     /note="MIR repeat: matches 79. .252 of consensus"
     repeat_region   27684. .28115
                     /note="MLT1J repeat: matches 36. .472 of consensus"
     repeat_region   28295. .28518
                     /note="MIR repeat: matches 29. .250 of consensus"
     repeat_region   29676. .29957
                     /note="L2 repeat: matches 2252. .2520 of consensus"
     repeat_region   30268. .30557
                     /note="AluSx repeat: matches 1. .293 of consensus"
     repeat_region   30558. .30693
                     /note="17 copies 8 mer tatatata 84% conserved"
     repeat_region   30560. .30695
                     /note="34 copies 4 mer tata 85% conserved"
     repeat_region   31523. .31813
                     /note="AluSq repeat: matches 1. .291 of consensus"
     repeat_region   31917. .31960
                     /note="22 copies 2 mer ac 93% conserved"
     repeat_region   31918. .31961
                     /note="11 copies 4 mer caca 95% conserved"
     repeat_region   31920. .31959
                     /note="5 copies 8 mer cacacaca 97% conserved"
     repeat_region   32433. .32745
                     /note="L1MB2 repeat: matches 5860. .6164 of consensus"
     repeat_region   32820. .33044
                     /note="MER1B repeat: matches 125. .337 of consensus"
     repeat_region   33045. .33306
                     /note="L1MB7 repeat: matches 5628. .5877 of consensus"
     repeat_region   33333. .33788
                     /note="L1MB7 repeat: matches 5729. .6173 of consensus"
     repeat_region   34018. .34081
                     /note="32 copies 2 mer ta 75% conserved"
     repeat_region   34024. .34079
                     /note="7 copies 8 mer tatatata 78% conserved"
     repeat_region   35097. .35280
                     /note="MER69 repeat: matches 74. .264 of consensus"
     repeat_region   36056. .36229
                     /note="L2 repeat: matches 2358. .2534 of consensus"
     repeat_region   36230. .36449
```

```
                /note="MER58A repeat: matches 1. .220 of consensus"
     repeat_region   36450. .36583
                /note="L2 repeat: matches 2533. .2666 of consensus"
     repeat_region   37373. .37723
                /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   38181. .38399
                /note="L2 repeat: matches 2457. .2656 of consensus"
     repeat_region   38402. .38745
                /note="MER1B repeat: matches 1. .337 of consensus"
     repeat_region   38884. .39041
                /note="MER45 repeat: matches 5. .178 of consensus"

  Query Match            72.8%;  Score 18.2;  DB 9;  Length 72289;
  Best Local Similarity  87.0%;  Pred. No. 3.8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcagcacgttg 25
          || |||||||||| ||||||| |
Db  38249 GTGAGGTGCAGGTAAGCACGTGG 38271


RESULT  26
AC003096
LOCUS       AC003096    72919 bp    DNA             PLN       05-APR-2000
DEFINITION  Arabidopsis thaliana chromosome II section 191 of 255 of the
            complete sequence. Sequence from clones T31E10, T29F13.
ACCESSION   AC003096 AE002093
VERSION     AC003096.2  GI:6598389
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 72919)
  AUTHORS   Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
            Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
            Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
            Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
            Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
            Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
            Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
            Venter,J.C.
  TITLE     Sequence and analysis of chromosome 2 of the plant Arabidopsis
            thaliana
  JOURNAL   Nature 402 (6763), 761-768 (1999)
  MEDLINE   20083487
   PUBMED   10617197
REFERENCE   2  (bases 1 to 72919)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Dec 17, 1999 this sequence version replaced gi:3132469.
            The sequence and annotation of chromosome 2 were merged from those
            of the individual clones on this chromosome after removing
            overlaps. For detailed information, please see the TIGR web site
            (http://www.tigr.org/tdb/at/at.html).

            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL
            (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,
            University of Washington), Genscan (Chris Burge,
            http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene
            (http://www.cbs.dtu.dk/services/NetGene2/), searches of the
            complete sequence against a peptide database and plant EST
            databases at TIGR, and manual curations based on those analyses.
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as 'unknown' proteins.
            Genes without protein or EST similarity, that are predicted by two
            or more gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are
            numbered from the top to bottom of the chromosome.

            We thank the CSHL/WashU/ABI consortium for sequencing BAC clones
            F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone
            F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
            and Satoshi Tabata for helpful assistance. In addition, we would
            like to thank the TIGR Bioinformatics Department, especially Lixin
            Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
            Peterson, Michael Holmes, and Delwood Richardson for software and
            database support.

            This work was supported by the National Science Foundation,
            Department of Energy and the US Department of Agriculture.

            Address all correspondence to:  at@tigr.org.
FEATURES            Location/Qualifiers
     source         1. .72919
                    /organism="Arabidopsis thaliana"
                    /cultivar="Columbia"
                    /db_xref="taxon:3702"
                    /chromosome="II"
     misc_feature   complement(<1. .5635)
                    /note="Sequence from clone T31E10"
     repeat_region  complement(234. .349)
                    /rpt_family="(CAAAA)n"
     repeat_region  464. .491
                    /rpt_family="POLY_A"
     repeat_region  1667. .1709
                    /rpt_family="(TA)n"
     mRNA           join(5636. .6518,6612. .6790,6877. .6916,7039. .7195,
                    7361. .7475,7554. .7634,8023. .8047,8274. .8330,8425. .8490,
                    8771. .8845,9250. .9357,9777. .9911,10048. .10154,
                    10232. .10325,10410. .10577,10752. .10892,10993. .11193,
                    11414. .11512,11613. .11765,11888. .12285,12468. .12648,
                    12857. .13180,13384. .14262,14338. .14583,14677. .15091,
                    15212. .15488)
                    /gene="At2g34660"
     gene           5636. .15488
                    /gene="At2g34660"
                    /note="T29F13.13; identical to GB:AF014960; transports
                    glutathione conjugates"
     misc_feature   complement(5636. .>72919)
                    /note="Sequence from clone T29F13"
     CDS            join(6189. .6518,6612. .6790,6877. .6916,7039. .7195,
                    7361. .7475,7554. .7634,8023. .8047,8274. .8330,8425. .8490,
                    8771. .8845,9250. .9357,9777. .9911,10048. .10154,
                    10232. .10325,10410. .10577,10752. .10892,10993. .11193,
                    11414. .11512,11613. .11765,11888. .12285,12468. .12648,
                    12857. .13180,13384. .14262,14338. .14583,14677. .15091,
                    15212. .15309)
                    /gene="At2g34660"
                    /codon_start=1
                    /product="ABC transporter (AtMRP2)"
                    /protein_id="AAC16268.1"
                    /db_xref="GI:3132479"
                    /translation="MGFEFIEWYCKPVPNGVWTKQVANAFGAYTPCATDSFVLGISQL
                    VLLVLCLYRIWLALKDHKVERFCLRSRLYNYFLALLAAYATAEPLFRLIMGISVLDFD
                    GPGLPPFEAFGLGVKAFAWGAVMVMILMETKIYIRELRWYVRFAVIYALVGDMVLLNL
                    VLSVKEYYSSYVLYLYTSEVGAQVLFGILLFMHLPNLDTYPGYMPVRSETVDDYEYEE
                    ISDGQQICPEKHANIFDKIFFSWMNPLMTLGSKRPLTEKDVWYLDTWDQTETLFTSFQ
                    HSWDKELQKPQPWLLRALNNSLGGRFWWGGFWKIGNDCSQFVGPLLLNQLLKSMQEDA
                    PAWMGYIYAFSIFVGVVFGVLCEAQYFQNVMRVGYRLRSALIAAVFRKSLRLTNEGRR
                    KFQTGKITNLMTTDAESLQQICQSLHTMWSAPFRIIIALILLYQQLGVASLIGALLLV
                    LMFPLQTVIISKMQKLTKEGLQRTDKRIGLMNEVLAAMDTVKCYAWENSFQSKVQTVR
                    DDELSWFRKSQLLGALNMFILNSIPVLVTIVSFGVFTLLGGDLTPARAFTSLSLFAVL
                    RFPLFMLPNIITQVVNANVSLKRLEEVLATEERILLPNPPIEPGEPAISIRNGYFSWD
                    SKGDRPTLSNINLDVPLGSLVAVVGSTGEGKTSLISAILGELPATSDAIVTLRGSVAY
                    VPQVSWIFNATVRDNILFGSPFDREKYERAIDVTSLKHDLELLPGGDLTEIGERGVNI
```

```
                SGGQKQRVSMARAVYSNSDVYIFDDPLSALDAHVGQQVFEKCIKRELGQKTRVLVTNQ
                LHFLSQVDRIVLVHEGTVKEEGTYEELSSNGPLFQRLMENAGKVEEYSEENGEAEADQ
                TAEQPVANGNTNGLQMDGSDDKKSKEGNKKGGKSVLIKQEERETGVVSWRVLKRYQDA
                LGGAWVVMMLLLCYVLTEVFRVTSSTWLSEWTDAGTPKSHGPLFYNLIYALLSFGQVL
                VTLTNSYWLIMSSLYAAKKLHDNMLHSILRAPMSFFHTNPLGRIINRFAKDLGDIDRT
                VAVFVNMFMGQVSQLLSTVVLIGIVSTLSLWAIMPLLVLFYGAYLYYQNTAREVKRMD
                SISRSPVYAQFGEALNGLSTIRAYKAYDRMADINGRSMDNNIRFTLVNMGANRWLGIR
                LETLGGLMIWLTASFAVMQNGRAENQQAFASTMGLLLSYALNITSLLTGVLRLASLAE
                NSLNAVERVGNYIEIPPEAPPVIENNRPPPGWPSSGSIKFEDVVLRYRPQLPPVLHGV
                SFFIHPTDKVGIVGRTGAGKSSLLNALFRIVEVEKGRILIDDCDVGKFGLMDLRKVLG
                IIPQSPVLFSGTVRFNLDPFGEHNDADLWESLERAHLKDTIRRNPLGLDAEVSEAGEN
                FSVGQRQLLSLSRALLRRSKILVLDEATAAVDVRTDALIQKTIREEFKSCTMLIIAHR
                LNTIIDCDKILVLDSGRVQEFSSPENLLSNEGSSFSKMVQSTGAANAEYLRSLVLDNK
                RAKDDSHHLQGQRKWLASSRWAAAAQFALAASLTSSHNDLQSLEIEDDSSILKRTNDA
                VVTLRSVLEGKHDKEIAESLEEHNISREGWLSSLYRMVEGLAVMSRLARNRMQQPDYN
                FEGNTFDWDNVEM"
misc_feature    complement(13668. .14027)
                /note="molecular marker ve016"
mRNA            complement(join(<15663. .16061,16251. .16560,16668. .16891,
                16973. .17154,17274. .>18153))
                /gene="At2g34670"
gene            complement(<15663. .>18153)
                /gene="At2g34670"
                /note="T29F13.12"
CDS             complement(join(15663. .16061,16251. .16560,16668. .16891,
                16973. .17154,17274. .18153))
                /gene="At2g34670"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="AAC16267.1"
                /db_xref="GI:3132478"
                /translation="MGCAASRIDNEEKVLVCRQRKRLMKKLLGFRGEFADAQLAYLRA
                LRNTGVTLRQFTESETLELENTSYGLSLPLPPSPPPTLPPSPPPPPPPFSPDLRNPETS
                HDLADEEEEGENDGGNDGSGAAPPPPLPNSWNIWNPFESLELHSHPNGDNVVTQVELK
                KKQQTQQAEEEDWAETKSQFEEEDEQQEAGGTCLDLSVHQIEAVSGCNMKKPRRLKFK
                LGEVMDGNSSMTSCSGKDLEKTHVTDCR1RRTLEGIIRELDDYFLKASGCEKEIAVIV
                DINSRDTVDPFRYQETRRKRSSSAKVFSALSWSWSSKSLQLGKDATTSGTVEPCRPGA
                HCSTLEKLYTAEKKLYQLVRNKEIAKVEHERKSALLQKQDGETYDLSKMEKARLSLES
                LETEIQRLEDSITTTRSCLLNLINDELYPQLVALTSGLAQMWKTMLKCHQVQTHISQQ
                LNHLPDYPSIDLSSEYKRQAVNELETEVTCWYNSFCKLVNSQREYVKTLCTWIQLTDR
                LSNEDNQRSSLPVAARKLCKEWQLEYNLRRKCNKLERRLEKELISLAEIERRLEGILA
                MEEEEVSSTSLGSKHPLSIKQAKIEALRKRVDIEKTKYLNSVEVSKRMTLDNLKSSLP
                NVFQMLTALANVFANGFESVNGQTGTDVSDTSQHSDESQP"
repeat_region   complement(17583. .17677)
                /rpt_family="(GAA)n"
repeat_region   17882. .17937
                /rpt_family="(GGA)n"
repeat_region   18888. .18927
                /rpt_family="POLY_A"
mRNA            complement(join(19728. .20045,20129. .20260,20346. .20435,
                20465. .20523,20774. .20905,21067. .21222,21309. .21490,
                21666. .21849,21949. .22022,22320. .22524,22781. .22902,
                22992. .23187,23344. .23463,23557. .23630,23712. .23796,
                23837. .24080,24191. .24309,24596. .24772,25433. .25550,
                25630. .25814,26314. .26414,26518. .26653,27022. .27197,
                27322. .27467,27751. .27983,28270. .28371,28832. .28931,
                29231. .29296,29387. .29449,29582. .29647,29732. .29803,
                29894. .29959,30240. .30311,30410. .>31222))
                /gene="At2g34680"
gene            complement(19728. .>31222)
                /gene="At2g34680"
                /note="T29F13.11"
CDS             complement(join(19869. .20045,20129. .20260,20346. .20435,
                20465. .20523,20774. .20905,21067. .21222,21309. .21490,
                21666. .21849,21949. .22022,22320. .22524,22781. .22902,
                22992. .23187,23344. .23463,23557. .23630,23712. .23796,
                23837. .24080,24191. .24309,24596. .24772,25433. .25550,
                25630. .25814,26314. .26414,26518. .26653,27022. .27197,
                27322. .27467,27751. .27983,28270. .28371,28832. .28931,
                29231. .29296,29387. .29449,29582. .29647,29732. .29803,
                29894. .29959,30240. .30311,30410. .31222))
                /gene="At2g34680"
                /note="unknown protein"
                /codon_start=1
                /protein_id="AAC16266.1"
                /db_xref="GI:3132477"
                /translation="MEEVAAKVEEETVETNVDAVKEDNATIANESRSPESVSAVSVVS

  Query Match             72.8%;  Score 18.2;  DB 8;  Length 72919;
  Best Local Similarity   87.0%;  Pred. No. 3.8e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgt 23
           || ||| ||||| ||||||||||
Db   55600 TCCTCAAGTGCATGTCAGCACGT 55622


RESULT  27
ATT20O10
LOCUS       ATT20O10    83122 bp    DNA             PLN       13-APR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 3, BAC clone T20O10.
ACCESSION   AL163816
VERSION     AL163816.1  GI:7573419
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 83122)
  AUTHORS   Obermaier,B., Ottenwaelder,B., Duchemin,D., Zeitler,K., Mewes,H.W.,
            Rudd,S., Lemcke,K., Mayer,K.F.X., Quetier,F. and Salanoubat,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 83122)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
            d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
            Gaston Cremieux, BP191, 91006 Evry Cedex, France;
            http://www.genoscope.cns.fr
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .83122
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="3"
     gene            547. .3272
                     /gene="T20O10_10"
     gene            complement(join(547. .627,715. .798,874. .921,1019. .1076,
                     1175. .1307,1385. .1457,1558. .1709,1847. .1912,2119. .2172,
                     2319. .2370,2485. .2588,2679. .2736,2811. .2891,2971. .3027,
                     3114. .3272))
                     /gene="T20O10_10"
     exon            complement(547. .627)
                     /gene="T20O10_10"
                     /number=1
     CDS             complement(join(547. .627,715. .798,874. .921,1019. .1076,
                     1175. .1307,1385. .1457,1558. .1709,1847. .1912,2119. .2172,
                     2319. .2370,2485. .2588,2679. .2736,2811. .2891,2971. .3027,
                     3114. .3272))
                     /gene="T20O10_10"
                     /note="strong similarity to translation releasing factor
                     RF-1, Synechocystis sp., PIR:S76914
                     Contains Prokaryotic-type class I peptide chain release
                     factors signature AA280-296"
                     /codon_start=1
                     /product="translation releasing factor RF-1-like protein"
                     /protein_id="CAB87736.1"
                     /db_xref="GI:7573420"
                     /translation="MNSSMTTSPTARAFLHLHSSPFTLRSRKSPSGVCLVSWPPSRSR
```

```
                RTRRLVCMAEPYLIRKMESVEKTWKELSVKLADPDVVSNQSEYQKLAQSMSELDEVVT
                VFRRFKDCEKQLLESKVLAKEAGDDEDMAEMIGSEINSLTKEIEELEKQLKVLLLPSD
                PLDARNILLEVRAGTGGDEAAIWTGDLVGADVSEVQRAQLLEVFNGFMLRGYKTCVME
                IKGNRVYSKLKYESGVHRVQRVPQTETQGRVHTSTATVAIMPEADEVEVVIDPKDIEL
                TSARSGGAGGQNVNKVETAIDLFHKPSGIRIFCTEERTQIRNKARAFQLLRAKLYEIK
                VREQQEKIRNERKSQVGTGARSEKIRTYNYKDSRVTDHRLKMNFALTTFLDGALEDAV
                QACAALEQKELMEELSESVAASSATSG"
intron          complement(628. .714)
                /number=1
exon            complement(715. .798)
                /gene="T20O10_10"
                /number=2
intron          complement(799. .873)
                /number=2
exon            complement(874. .921)
                /gene="T20O10_10"
                /number=3
intron          complement(922. .1018)
                /number=3
exon            complement(1019. .1076)
                /gene="T20O10_10"
                /number=4
intron          complement(1077. .1174)
                /number=4
exon            complement(1175. .1307)
                /gene="T20O10_10"
                /number=5
intron          complement(1308. .1384)
                /number=5
exon            complement(1385. .1457)
                /gene="T20O10_10"
                /number=6
intron          complement(1458. .1557)
                /number=6
exon            complement(1558. .1709)
                /gene="T20O10_10"
                /number=7
intron          complement(1710. .1846)
                /number=7
exon            complement(1847. .1912)
                /gene="T20O10_10"
                /number=8
intron          complement(1913. .2118)
                /number=8
exon            complement(2119. .2172)
                /gene="T20O10_10"
                /number=9
intron          complement(2173. .2318)
                /number=9
exon            complement(2319. .2370)
                /gene="T20O10_10"
                /number=10
intron          complement(2371. .2484)
                /number=10
exon            complement(2485. .2588)
                /gene="T20O10_10"
                /number=11
intron          complement(2589. .2678)
                /number=11
exon            complement(2679. .2736)
                /gene="T20O10_10"
                /number=12
intron          complement(2737. .2810)
                /number=12
exon            complement(2811. .2891)
                /gene="T20O10_10"
                /number=13
intron          complement(2892. .2970)
                /number=13
exon            complement(2971. .3027)
                /gene="T20O10_10"
                /number=14
intron          complement(3028. .3113)
                /number=14
exon            complement(3114. .3272)
                /gene="T20O10_10"
                /number=15
gene            3603. .4057
                /gene="T20O10_20"
exon            3603. .3652
                /gene="T20O10_20"
                /number=1
CDS             join(3603. .3652,3865. .4057)
                /gene="T20O10_20"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB87737.1"
                /db_xref="GI:7573421"
                /translation="MSIYRIARALPFSGILRQLEKEAETVINVLQPGPLGIIEHKFSA
                QEIREAKATVSRAVDSWRRHSEVEHANGLLKDYIYK"
intron          3653. .3864
                /gene="T20O10_20"
                /number=1
exon            3865. .4057
                /gene="T20O10_20"
                /number=2
gene            complement(4424. .4732)
                /gene="T20O10_30"
gene            4424. .4732
                /gene="T20O10_30"
CDS             complement(4424. .4732)
                /gene="T20O10_30"
                /note="similarity to glutaredoxin, castor bean,
                PIR:S54825"
                /codon_start=1
                /product="glutaredoxin-like protein"
                /protein_id="CAB87738.1"
                /db_xref="GI:7573422"
                /translation="MESVRSLVEDKPVVIFSKSSCCMSHSIQTLISGFGAKMTVYELD
                QFSNGQEIEKALVQMGCKPSVPAVFIGQQFIGGANQVMTLQVKNQLAAMLRRAGAIWV
                "
exon            complement(4424. .4732)
                /gene="T20O10_30"
                /number=1
gene            5992. .7131
                /gene="T20O10_40"
gene            complement(join(5992. .6883,7025. .7131))
                /gene="T20O10_40"
exon            complement(5992. .6883)
                /gene="T20O10_40"
                /number=1
CDS             complement(join(5992. .6883,7025. .7131))
                /gene="T20O10_40"
                /note="similarity to CGI-77 protein, Homo sapiens,
                EMBL:AF151836"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB87739.1"
                /db_xref="GI:7573423"
                /translation="MADNKTSEGIVSEETPMDASQEQQHETVEEMLARHRQEIKQLQN
                KETEMKKAAAKGSKAEQKAKKKQVEDDISKLSTKLKDKQLKELASQGFSSSSSNNIAK
                DETTEKKGDIDTLVRAIAGVSVTAQQEHSKPSKSVKRREKRAKEEADREQRIKEEQSH
                VKSDRMVENAKLEKKLKPLGLTVSEIKPDGHCLYRAVENQLANRSGGASPYTYQNLRE
                MAASYMREHKTDFLPFFLSETEGDSNSGSAEERFEKYCREVESTAAWGSQLELGALTH
                CLRKHIKVYSGSFPDVEMGKEYRSGDDSSLMLSYHRHAFGLGEHYNSVVLVNNITG"
intron          complement(6884. .7024)
                /number=1
exon            complement(7025. .7131)
                /gene="T20O10_40"
                /number=2
exon            9189. .9500
                /gene="T20O10_50"
                /number=1

Query Match             72.8%;  Score 18.2;  DB 8;  Length 83122;
Best Local Similarity   87.0%;  Pred. No. 3.7e+02;
```

```
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgtt 24
           ||||||||  ||  ||||||||  ||
Db   65476 CGTCAGGTCCATGTCAGCACCTT 65498


RESULT  28
AC022591/c
LOCUS       AC022591    93330 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens clone RP11-21J18, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC022591
VERSION     AC022591.2  GI:9124011
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 93330)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-21J18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 93330)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6910737.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4035
                Center clone name: 21_J_18
            --------------.
            * NOTE: This record contains 91 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      894: contig of 894 bp in length
            *      895 994: gap of      100 bp
            *      995     1895: contig of 901 bp in length
            *     1896 1995: gap of      100 bp
            *     1996     2929: contig of 934 bp in length
            *     2930 3029: gap of      100 bp
            *     3030     3960: contig of 931 bp in length
            *     3961 4060: gap of      100 bp
            *     4061     5031: contig of 971 bp in length
            *     5032 5131: gap of      100 bp
            *     5132     6053: contig of 922 bp in length
            *     6054 6153: gap of      100 bp
            *     6154     7062: contig of 909 bp in length
            *     7063 7162: gap of      100 bp
            *     7163     8056: contig of 894 bp in length
            *     8057 8156: gap of      100 bp
            *     8157     9085: contig of 929 bp in length
            *     9086 9185: gap of      100 bp
            *     9186    10128: contig of 943 bp in length
            *    10129 10228: gap of      100 bp
            *    10229    11163: contig of 935 bp in length
            *    11164 11263: gap of      100 bp
            *    11264    12165: contig of 902 bp in length
            *    12166 12265: gap of      100 bp
            *    12266    13196: contig of 931 bp in length
            *    13197 13296: gap of      100 bp
            *    13297    14198: contig of 902 bp in length
            *    14199 14298: gap of      100 bp
            *    14299    15211: contig of 913 bp in length
            *    15212 15311: gap of      100 bp
            *    15312    16213: contig of 902 bp in length
            *    16214 16313: gap of      100 bp
            *    16314    17266: contig of 953 bp in length
            *    17267 17366: gap of      100 bp
            *    17367    18307: contig of 941 bp in length
            *    18308 18407: gap of      100 bp
            *    18408    19326: contig of 919 bp in length
            *    19327 19426: gap of      100 bp
            *    19427    20356: contig of 930 bp in length
            *    20357 20456: gap of      100 bp
            *    20457    21367: contig of 911 bp in length
            *    21368 21467: gap of      100 bp
            *    21468    22359: contig of 892 bp in length
            *    22360 22459: gap of      100 bp
            *    22460    23377: contig of 918 bp in length
            *    23378 23477: gap of      100 bp
            *    23478    24384: contig of 907 bp in length
            *    24385 24484: gap of      100 bp
            *    24485    25418: contig of 934 bp in length
            *    25419 25518: gap of      100 bp
            *    25519    26448: contig of 930 bp in length
            *    26449 26548: gap of      100 bp
            *    26549    27508: contig of 960 bp in length
            *    27509 27608: gap of      100 bp
            *    27609    28530: contig of 922 bp in length
            *    28531 28630: gap of      100 bp
            *    28631    29542: contig of 912 bp in length
            *    29543 29642: gap of      100 bp
            *    29643    30587: contig of 945 bp in length
            *    30588 30687: gap of      100 bp
            *    30688    31615: contig of 928 bp in length
            *    31616 31715: gap of      100 bp
            *    31716    32650: contig of 935 bp in length
            *    32651 32750: gap of      100 bp
            *    32751    33680: contig of 930 bp in length
            *    33681 33780: gap of      100 bp
            *    33781    34707: contig of 927 bp in length
            *    34708 34807: gap of      100 bp
            *    34808    35781: contig of 974 bp in length
            *    35782 35881: gap of      100 bp
            *    35882    36784: contig of 903 bp in length
            *    36785 36884: gap of      100 bp
            *    36885    37813: contig of 929 bp in length
            *    37814 37913: gap of      100 bp
            *    37914    38865: contig of 952 bp in length
            *    38866 38965: gap of      100 bp
            *    38966    39920: contig of 955 bp in length
            *    39921 40020: gap of      100 bp
            *    40021    40937: contig of 917 bp in length
```

```
*    40938 41037: gap of      100 bp
*    41038    41980: contig of 943 bp in length
*    41981 42080: gap of      100 bp
*    42081    42997: contig of 917 bp in length
*    42998 43097: gap of      100 bp
*    43098    43997: contig of 900 bp in length
*    43998 44097: gap of      100 bp
*    44098    45007: contig of 910 bp in length
*    45008 45107: gap of      100 bp
*    45108    46068: contig of 961 bp in length
*    46069 46168: gap of      100 bp
*    46169    47090: contig of 922 bp in length
*    47091 47190: gap of      100 bp
*    47191    48084: contig of 894 bp in length
*    48085 48184: gap of      100 bp
*    48185    49107: contig of 923 bp in length
*    49108 49207: gap of      100 bp
*    49208    50167: contig of 960 bp in length
*    50168 50267: gap of      100 bp
*    50268    51211: contig of 944 bp in length
*    51212 51311: gap of      100 bp
*    51312    52258: contig of 947 bp in length
*    52259 52358: gap of      100 bp
*    52359    53260: contig of 902 bp in length
*    53261 53360: gap of      100 bp
*    53361    54277: contig of 917 bp in length
*    54278 54377: gap of      100 bp
*    54378    55306: contig of 929 bp in length
*    55307 55406: gap of      100 bp
*    55407    56360: contig of 954 bp in length
*    56361 56460: gap of      100 bp
*    56461    57356: contig of 896 bp in length
*    57357 57456: gap of      100 bp
*    57457    58417: contig of 961 bp in length
*    58418 58517: gap of      100 bp
*    58518    59425: contig of 908 bp in length
*    59426 59525: gap of      100 bp
*    59526    60442: contig of 917 bp in length
*    60443 60542: gap of      100 bp
*    60543    61447: contig of 905 bp in length
*    61448 61547: gap of      100 bp
*    61548    62497: contig of 950 bp in length
*    62498 62597: gap of      100 bp
*    62598    63505: contig of 908 bp in length
*    63506 63605: gap of      100 bp
*    63606    64553: contig of 948 bp in length
*    64554 64653: gap of      100 bp
*    64654    65605: contig of 952 bp in length
*    65606 65705: gap of      100 bp
*    65706    66663: contig of 958 bp in length
*    66664 66763: gap of      100 bp
*    66764    67658: contig of 895 bp in length
*    67659 67758: gap of      100 bp
*    67759    68638: contig of 880 bp in length
*    68639 68738: gap of      100 bp
*    68739    69643: contig of 905 bp in length
*    69644 69743: gap of      100 bp
*    69744    70691: contig of 948 bp in length
*    70692 70791: gap of      100 bp
*    70792    71697: contig of 906 bp in length
*    71698 71797: gap of      100 bp
*    71798    72708: contig of 911 bp in length
*    72709 72808: gap of      100 bp

  Query Match             72.8%;  Score 18.2;  DB 2;  Length 93330;
  Best Local Similarity   87.0%;  Pred. No. 3.7e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           || |||||||||| ||||||| |
Db   47869 GTGAGGTGCAGGTAAGCACGTGG 47847
```

```
RESULT  29
AF312913/c
LOCUS       AF312913   122951 bp    DNA             PRI       13-JUN-2001
DEFINITION  Homo sapiens chromosome 20 clone h119, complete sequence.
ACCESSION   AF312913
VERSION     AF312913.1  GI:11094029
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 122951)
  AUTHORS   Collins,C., Volik,S., Kowbel,D., Ginzinger,D., Ylstra,B.,
            Cloutier,T., Hawkins,T., Predki,P., Martin,C., Wernick,M.,
            Kuo,W.L., Alberts,A. and Gray,J.W.
  TITLE     Comprehensive genome sequence analysis of a breast cancer amplicon
  JOURNAL   Genome Res. 11 (6), 1034-1042 (2001)
  MEDLINE   21275464
   PUBMED   11381030
REFERENCE   2  (bases 1 to 122951)
  AUTHORS   Volik,S.V., Collins,C.C., Gray,J.W., Stultz,K. and Martin,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2000) Cancer Genetics, UCSF Comprehensive Cancer
            Center, 2340 Sutter St, Rm. S151, San Francisco, CA 94115, USA
FEATURES             Location/Qualifiers
     source          1. .122951
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="h119"
     misc_feature    146. .2513
                     /note="overlaps Homo sapiens chromosome 20 BAC clone 104
                     (LBNL H109) deposited in GenBank Accession Number
                     AC004499.1"
     misc_feature    complement(246. .49074)
                     /note="overlaps Homo sapiens chromosome 20 clone
                     RP11-15M15 deposited in GenBank Accession Number
                     AL391097.11"
     repeat_region   complement(2999. .3090)
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   complement(4914. .4955)
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   5266. .5853
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   6314. .6595
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   6803. .6966
                     /rpt_family="MER20"
                     /rpt_type=dispersed
     repeat_region   9285. .9382
                     /rpt_family="L1"
                     /rpt_type=dispersed
     repeat_region   12366. .12727
                     /rpt_family="THE1"
                     /rpt_type=dispersed
     repeat_region   complement(14379. .14604)
                     /rpt_family="7SK"
                     /rpt_type=dispersed
     repeat_region   15110. .15225
                     /rpt_family="MER20"
                     /rpt_type=dispersed
     repeat_region   complement(15343. .15651)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   15798. .16091
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   16422. .16619
                     /rpt_family="MER20"
```

```
                     /rpt_type=dispersed
     repeat_region   complement(20205. .20484)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     misc_feature    complement(20496. .87806)
                     /note="similar to clone RP5-911I5 on chromosome
                     20q13.11-13.2 deposited in GenBank Accession Number
                     AL050316.15"
     repeat_region   complement(23398. .23719)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   26809. .26864
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   27469. .27714
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   27968. .28297
                     /rpt_family="MER1"
                     /rpt_type=dispersed
     misc_feature    complement(29317. .31331)
                     /note="similar to Homo sapiens chromosome 5 clone CTB-35K5
                     deposited in GenBank Accession Number AC011401.7"
     repeat_region   complement(31334. .31608)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(36709. .36837)
                     /rpt_family="MER20"
                     /rpt_type=dispersed
     repeat_region   40808. .40990
                     /rpt_family="MLT1"
                     /rpt_type=dispersed
     repeat_region   41982. .42226
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(44154. .44253)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   45474. .46247
                     /rpt_family="Tigger1"
                     /rpt_type=dispersed
     repeat_region   complement(46536. .46612)
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     STS             complement(48595. .49074)
                     /standard_name="SHGC-77795"
                     /evidence=not_experimental
                     /db_xref="dbSTS:G49984"
     STS             49507. .49923
                     /standard_name="SHGC-78826"
                     /evidence=not_experimental
                     /db_xref="dbSTS:G48638"
     STS             49520. .50008
                     /standard_name="SHGC-78862"
                     /db_xref="dbSTS:G48652"
     STS             49551. .49924
                     /standard_name="SHGC-82724"
                     /evidence=not_experimental
                     /db_xref="dbSTS:G48687"
     repeat_region   complement(50213. .50359)
                     /rpt_family="MER45"
                     /rpt_type=dispersed
     misc_feature    50367. .122949
                     /note="similar to Homo sapiens clone RP11-21J18 deposited
                     in GenBank Accession Number AC022591.2"
     repeat_region   50505. .50741
                     /rpt_family="MER1"
                     /rpt_type=dispersed
     repeat_region   50767. .50842
                     /rpt_family="MER1"
                     /rpt_type=dispersed
     repeat_region   complement(51523. .51873)
                     /rpt_family="THE1"
                     /rpt_type=dispersed
     repeat_region   complement(56228. .56300)
                     /rpt_family="MER1"
                     /rpt_type=dispersed
     repeat_region   complement(56315. .56511)
                     /rpt_family="MER1"
                     /rpt_type=dispersed
     repeat_region   complement(57433. .57713)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(58689. .58969)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   63803. .64058
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   64450. .64608
                     /rpt_family="MER46"
                     /rpt_type=dispersed
     repeat_region   complement(64504. .64688)
                     /rpt_family="MER46"
                     /rpt_type=dispersed
     repeat_region   complement(65294. .65499)
                     /rpt_family="MER20"
                     /rpt_type=dispersed
     repeat_region   67033. .67145
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   complement(68663. .68943)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(68995. .69186)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   69649. .69909
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(70559. .70669)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(74449. .74549)
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   74819. .75109
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(75767. .75983)
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   complement(77660. .77936)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(79568. .79670)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   80832. .80988

  Query Match            72.8%;  Score 18.2;  DB 9;  Length 122951;
  Best Local Similarity  87.0%;  Pred. No. 3.5e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           || |||||||||| ||||||| |
Db   50997 GTGAGGTGCAGGTAAGCACGTGG 50975


RESULT  30
AC004691/c
LOCUS       AC004691   141990 bp    DNA             PRI       03-FEB-2000
DEFINITION  Homo sapiens PAC clone RP4-740D2 from 7p14-p15, complete sequence.
ACCESSION   AC004691
VERSION     AC004691.1  GI:3135285
KEYWORDS    HTG.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 141990)
  AUTHORS   Dempsey,S., Lacy,M., Holmes,A. and Nguyen,C.
  TITLE     The sequence of Homo sapiens PAC clone RP4-740D2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 141990)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   3  (bases 1 to 141990)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     -------------- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu/gsc
              Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
              Center project name: H_DJ0740D02
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-4, prepared by
            Pieter de Jong and coworkers at Roswell Park Cancer Institute,
            using the method described by Ioannou et al., Nature Genetics
            6:84-9 (1994).  The library is from one male donor.  For further
            details, see http://bacpac.med.buffalo.edu/
            The clone is available from Genome Systems, Inc.
            (http://www.genomesystems.com).
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is RP5-877J2.  Actual start of
            this clone is at base position 1 of RP4-740D2; actual end is at
            141990 of RP4-740D2.

            This clone contains STS sWSS1731 (NID:g1113195).
FEATURES              Location/Qualifiers
     source           1. .141990
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="7"
                      /map="7p14-p15"
                      /clone="RP4-740D2"
                      /clone_lib="RPCI-4"
     repeat_region    751. .915
                      /rpt_family="MIR"
     repeat_region    1675. .1765
                      /rpt_family="MIR"
     repeat_region    1768. .1957
                      /rpt_family="L2"
     repeat_region    2057. .3427
                      /rpt_family="L2"
     repeat_region    3453. .4205
                      /rpt_family="MER21_g"
     repeat_region    4218. .4996
                      /rpt_family="L2"
     repeat_region    5208. .5290
                      /rpt_family="L2"
     repeat_region    5404. .5567
                      /rpt_family="Alu"
     repeat_region    10107. .10264
                      /rpt_family="L2"
     repeat_region    10631. .10678
                      /rpt_family="MaLR"
     repeat_region    10679. .10726
                      /rpt_family="Retroviral"
     repeat_region    10719. .10804
                      /rpt_family="(CATA)n"
     repeat_region    10815. .11116
                      /rpt_family="Alu"
     repeat_region    11118. .11424
                      /rpt_family="Retroviral"
     repeat_region    11446. .11824
                      /rpt_family="MaLR"
     repeat_region    11848. .11984
                      /rpt_family="T2_type"
     repeat_region    11986. .12048
                      /rpt_family="MaLR"
     repeat_region    14726. .15039
                      /rpt_family="Alu"
     repeat_region    16196. .17566
                      /rpt_family="MER1_type"
     repeat_region    18844. .19139
                      /rpt_family="Alu"
     repeat_region    19186. .19256
                      /rpt_family="MIR"
     repeat_region    20185. .20389
                      /rpt_family="MER1_type"
     repeat_region    20582. .21615
                      /rpt_family="L1"
     repeat_region    21811. .22062
                      /rpt_family="Alu"
     repeat_region    22336. .22400
                      /rpt_family="MIR"
     repeat_region    23870. .23933
                      /rpt_family="MIR"
     repeat_region    23985. .24487
                      /rpt_family="MaLR"
     repeat_region    25373. .25574
                      /rpt_family="MER1_type?"
     repeat_region    25583. .26294
                      /rpt_family="L2"
     repeat_region    26438. .26697
                      /rpt_family="MER2_type"
     repeat_region    26712. .26847
                      /rpt_family="L2"
     repeat_region    26893. .27192
                      /rpt_family="Alu"
     repeat_region    27979. .28151
                      /rpt_family="MER1_type"
     repeat_region    28152. .28414
                      /rpt_family="Retroviral"
     repeat_region    31181. .31300
                      /rpt_family="(CCAA)n"
     repeat_region    33019. .33151
                      /rpt_family="MIR"
     repeat_region    34032. .34332
                      /rpt_family="L2"
```

```
     repeat_region   35060. .35102
                     /rpt_family="L2"
     repeat_region   35922. .36389
                     /rpt_family="Retroviral"
     repeat_region   36390. .37203
                     /rpt_family="Retroviral"
     repeat_region   37283. .37402
                     /rpt_family="(GA)n"
     repeat_region   37404. .37951
                     /rpt_family="Retroviral"
     repeat_region   38016. .38088
                     /rpt_family="MER4-group"
     repeat_region   38095. .38178
                     /rpt_family="Retroviral"
     repeat_region   38291. .38589
                     /rpt_family="Alu"
     repeat_region   38728. .39167
                     /rpt_family="MER4-group"
     repeat_region   39298. .39583
                     /rpt_family="Retroviral"
     repeat_region   39670. .39913
                     /rpt_family="Retroviral"
     repeat_region   39915. .40376
                     /rpt_family="Retroviral"
     repeat_region   40534. .40815
                     /rpt_family="L1"
     repeat_region   42486. .42671
                     /rpt_family="AcHobo"
     repeat_region   42860. .43006
                     /rpt_family="MIR"
     repeat_region   46335. .46765
                     /rpt_family="L1"
     repeat_region   46795. .47028
                     /rpt_family="MIR"
     repeat_region   47156. .47719
                     /rpt_family="L1"
     repeat_region   47779. .48269
                     /rpt_family="L1"
     repeat_region   48295. .48595
                     /rpt_family="Alu"
     repeat_region   48778. .49324
                     /rpt_family="L1"
     repeat_region   49647. .50115
                     /rpt_family="L1"
     repeat_region   50261. .50378
                     /rpt_family="L1"
     repeat_region   50415. .50545
                     /rpt_family="L1"
     repeat_region   50571. .50865

  Query Match             72.8%;  Score 18.2;  DB 9;  Length 141990;
  Best Local Similarity   87.0%;  Pred. No. 3.4e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           ||||||||||  ||| |||| |||
Db 103543 GTCAGGTGCATGTCTGCACATTG 103521


RESULT  31
AC007044
LOCUS       AC007044   180257 bp    DNA             HTG       10-SEP-2000
DEFINITION  Homo sapiens chromosome UNK clone CTA-435J10, WORKING DRAFT
            SEQUENCE, 1 unordered pieces.
ACCESSION   AC007044
VERSION     AC007044.3  GI:10048128
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180257)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180257)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 10, 2000 this sequence version replaced gi:9838354.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_RG435J10
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; %
            Sequencing vector: plasmid; %
            Chemistry: Dye-primer ET; % of reads
            Chemistry: Dye-terminator Big Dye; % of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 180641 bases at least Q40
            Consensus quality: 181075 bases at least Q30
            Consensus quality: 181328 bases at least Q20
            Insert size: 159000; agarose-fp
            Insert size: 181700; sum-of-contigs
            Quality coverage:  in Q20 bases; agarose-fp
            Quality coverage:  in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   180257: contig of 180257 bp in length.
FEATURES              Location/Qualifiers
     source           1. .180257
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="UNK"
                      /clone="CTA-435J10"
     misc_feature     1. .180257
                      /note="assembly_name:Contig36
                      clone_end:SP6
                      vector_side:right"
BASE COUNT    46943 a  45989 c  45264 g  42061 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 2;  Length 180257;
  Best Local Similarity   87.0%;  Pred. No. 3.3e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           ||||||||||  ||| |||| |||
Db 178754 GTCAGGTGCATGTCTGCACATTG 178776


RESULT  32
AC091668
LOCUS       AC091668   180484 bp    DNA             HTG       16-MAY-2001
DEFINITION  Pan troglodytes clone RP43-161K12, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC091668
VERSION     AC091668.1  GI:14091867
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      chimpanzee.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 180484)
  AUTHORS   Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S.,
            Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E., Lee-Lin,S.-Q.,
            Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B., Masiello,C.,
            Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
            Shevchenko,Y., Snyder,B., Stantripop,S., Thomas,J.W., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180484)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     -------------- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
                Center project name: cee
                Center clone name: 161K12
            -------------- Summary Statistics
                Sequencing vector: plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 179027 bases at least Q40
                Consensus quality: 179485 bases at least Q30
                Consensus quality: 179679 bases at least Q20
                Insert size: 164000; agarose-fp
                Insert size: 180184; sum-of-contigs
                Quality coverage: 13.65x in Q20 bases; agarose-fp
                Quality coverage: 12.43x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     8312: contig of 8312 bp in length
            *     8313     8412: gap of unknown length
            *     8413    19209: contig of 10797 bp in length
            *    19210    19309: gap of unknown length
            *    19310    50320: contig of 31011 bp in length
            *    50321    50420: gap of unknown length
            *    50421   180484: contig of 130064 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180484
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-161K12"
                     /clone_lib="RP43"
     misc_feature    1. .8312
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    8413. .19209
                     /note="assembly_fragment"
     misc_feature    19310. .50320
                     /note="assembly_fragment"
     misc_feature    50421. .180484
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    48110 a  46059 c  44263 g  41738 t    314 others
ORIGIN
```

```
  Query Match           72.8%;  Score 18.2;  DB 2;  Length 180484;
  Best Local Similarity  87.0%;  Pred. No. 3.3e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaugtcagcacgttg 25
          |||||||||| ||| |||| |||
Db 119199 GTCAGGTGCATGTCTGCACATTG 119221


RESULT  33
AC092408/c
LOCUS       AC092408   184042 bp    DNA             HTG       04-JUL-2001
DEFINITION  Papio cynocephalus anubis clone RP41-446H8, WORKING DRAFT SEQUENCE,
            5 unordered pieces.
ACCESSION   AC092408
VERSION     AC092408.1  GI:14595782
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      olive baboon.
  ORGANISM  Papio cynocephalus anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 184042)
  AUTHORS   Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L.,
            Granite,S., Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E.,
            Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B.,
            Masiello,C., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
            Pearson,R., Prasad,A., Shevchenko,Y., Snyder,B., Stantripop,S.,
            Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L.,
            Walker,M.A., Wetherby,K.D., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 184042)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     -------------- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
                Center project name: ccl
                Center clone name: 446H08
            -------------- Summary Statistics
                Sequencing vector: plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 182748 bases at least Q40
                Consensus quality: 183121 bases at least Q30
                Consensus quality: 183192 bases at least Q20
                Insert size: 142000; agarose-fp
                Insert size: 183642; sum-of-contigs
                Quality coverage: 13.90x in Q20 bases; agarose-fp
                Quality coverage: 10.75x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7384: contig of 7384 bp in length
            *     7385     7484: gap of unknown length
            *     7485    22267: contig of 14783 bp in length
            *    22268    22367: gap of unknown length
            *    22368    43038: contig of 20671 bp in length
            *    43039    43138: gap of unknown length
```

```
            *    43139    71914: contig of 28776 bp in length
            *    71915    72014: gap of unknown length
            *    72015   184042: contig of 112028 bp in length.
FEATURES             Location/Qualifiers
     source          1. .184042
                     /organism="Papio cynocephalus anubis"
                     /db_xref="taxon:9555"
                     /clone="RP41-446H8"
                     /clone_lib="RP41"
     misc_feature    1. .7384
                     /note="assembly_fragment"
     misc_feature    7485. .22267
                     /note="assembly_fragment"
     misc_feature    22368. .43038
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    43139. .71914
                     /note="assembly_fragment"
     misc_feature    72015. .184042
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    44694 a  44765 c  45729 g  48445 t    409 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 2;  Length 184042;
  Best Local Similarity   87.0%;  Pred. No. 3.3e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           ||||||||||  | | ||||||||
Db 127836 GTCAGGTGCATGCCTGCACGTTG 127814


RESULT  34
AC064835
ID   AC064835   standard; DNA; HTG; 202828 BP.
XX
AC   AC064835;
XX
SV   AC064835.3
XX
DT   26-APR-2000 (Rel. 63, Created)
DT   10-JUL-2000 (Rel. 64, Last updated, Version 6)
XX
DE   Homo sapiens chromosome 16 clone RP11-673P17, WORKING DRAFT SEQUENCE, 12
DE   unordered pieces.
XX
KW   HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-202828
RA   Waterston R.H.;
RT   "The sequence of Homo sapiens clone";
RL   Unpublished.
XX
RN   [2]
RP   1-202828
RA   Waterston R.H.;
RT   ;
RL   Submitted (22-APR-2000) to the EMBL/GenBank/DDBJ databases.
RL   Genome Sequencing Center, Washington University School of Medicine, 4444
RL   Forest Park Parkway, St. Louis, MO 63108, USA
XX
CC   On May 16, 2000 this sequence version replaced gi:7712280.
CC   ---------------- Genome Center ----------------
CC   Center: Washington University Genome Sequencing Center
CC   Center code: WUGSC
CC   Web site:http://genome.wustl.edu/gsc/index.shtml
CC   -------------- Project Information --------------
CC   Center project name: H_NH0673P17
CC   -------------- Summary Statistics  --------------
CC   Sequencing vector: M13; 100%
CC   Sequencing vector: plasmid; 0%
CC   Chemistry: Dye-primer ET; 100% of reads
CC   Chemistry: Dye-terminator Big Dye; 0% of reads
CC   Assembly program: Phrap; version 0.990319
CC   Consensus quality: 195107 bases at least Q40
CC   Consensus quality: 197538 bases at least Q30
CC   Consensus quality: 198979 bases at least Q20
CC   Insert size: 199000; agarose-fp
CC   Insert size: 201728; sum-of-contigs
CC   Quality coverage: 5.33 in Q20 bases; agarose-fp
CC   Quality coverage: 5.29 in Q20 bases; sum-of-contigs
CC   ------------------------------------------------.
CC   * NOTE: This is a 'working draft' sequence. It currently
CC   * consists of 12 contigs. The true order of the pieces
CC   * is not known and their order in this sequence record is
CC   * arbitrary. Gaps between the contigs are represented as
CC   * runs of N, but the exact sizes of the gaps are unknown.
CC   * This record will be updated with the finished sequence
CC   * as soon as it is available and the accession number will
CC   * be preserved.
CC   *        1     1109: contig of 1109 bp in length
CC   *     1110     1209: gap of unknown length
CC   *     1210     5935: contig of 4726 bp in length
CC   *     5936     6035: gap of unknown length
CC   *     6036    13020: contig of 6985 bp in length
CC   *    13021    13120: gap of unknown length
CC   *    13121    21028: contig of 7908 bp in length
CC   *    21029    21128: gap of unknown length
CC   *    21129    27413: contig of 6285 bp in length
CC   *    27414    27513: gap of unknown length
CC   *    27514    34551: contig of 7038 bp in length
CC   *    34552    34651: gap of unknown length
CC   *    34652    53143: contig of 18492 bp in length
CC   *    53144    53243: gap of unknown length
CC   *    53244    77201: contig of 23958 bp in length
CC   *    77202    77301: gap of unknown length
CC   *    77302   106156: contig of 28855 bp in length
CC   *   106157   106256: gap of unknown length
CC   *   106257   131644: contig of 25388 bp in length
CC   *   131645   131744: gap of unknown length
CC   *   131745   158698: contig of 26954 bp in length
CC   *   158699   158798: gap of unknown length
CC   *   158799   202828: contig of 44030 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .202828
FT                   /chromosome="16"
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
FT                   /clone="RP11-673P17"
FT   misc_feature    1. .1109
FT                   /note="assembly_name:Contig14"
FT   misc_feature    1210. .5935
FT                   /note="assembly_name:Contig18"
FT   misc_feature    6036. .13020
FT                   /note="assembly_name:Contig19"
FT   misc_feature    13121. .21028
FT                   /note="assembly_name:Contig20"
FT   misc_feature    21129. .27413
FT                   /note="assembly_name:Contig21 clone_end:T7
FT                   vector_side:right"
FT   misc_feature    27514. .34551
FT                   /note="assembly_name:Contig22"
FT   misc_feature    34652. .53143
FT                   /note="assembly_name:Contig23"
FT   misc_feature    53244. .77201
```

```
FT                   /note="assembly_name:Contig24"
FT   misc_feature    77302. .106156
FT                   /note="assembly_name:Contig25"
FT   misc_feature    106257. .131644
FT                   /note="assembly_name:Contig26 clone_end:SP6
FT                   vector_side:left"
FT   misc_feature    131745. .158698
FT                   /note="assembly_name:Contig27"
FT   misc_feature    158799. .202828
FT                   /note="assembly_name:Contig28"
XX
SQ   Sequence 202828 BP; 53011 A; 46987 C; 47342 G; 54379 T; 1109 other;


  Query Match             72.8%;  Score 18.2;  DB 33;  Length 202828;
  Best Local Similarity   87.0%;  Pred. No. 3.3e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           | ||||||||||||||||| || |
Db   46986 GCCAGGTGCAGGTCAGCAGGTGG 47008


RESULT  35
SME591789/c
LOCUS       SME591789  294800 bp    DNA             BCT       16-AUG-2001
DEFINITION  Sinorhizobium meliloti 1021 complete chromosome; segment 8/12.
ACCESSION   AL591789 AL591688
VERSION     AL591789.1  GI:15074950
KEYWORDS    .
SOURCE      Sinorhizobium meliloti.
  ORGANISM  Sinorhizobium meliloti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Sinorhizobium.
REFERENCE   1  (bases 1 to 294800)
  AUTHORS   Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
            Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
            Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
            Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ramsperger,U.,
            Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.
  TITLE     From the Cover: Analysis of the chromosome sequence of the legume
            symbiont Sinorhizobium meliloti strain 1021
  JOURNAL   Proceedings of the National Academy of Sciences of the United
            States of America. 98 (17), 9877-9882 (2001)
   PUBMED   11481430
REFERENCE   2  (bases 1 to 294800)
  AUTHORS   Gouzy,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
            EU Consortium
COMMENT     MELILO EU Consortium:
            Laboratoire de Biologie Moleculaire des Relations
            Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
            France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
            Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
            France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
            Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
            D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
            Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
            B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
            des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
            B-5030 Gembloux, Belgium.  E-mail:Jerome.Gouzy@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/meliloti.html.
FEATURES             Location/Qualifiers
     source          1. .294800
                     /organism="Sinorhizobium meliloti"
                     /strain="1021"
                     /db_xref="taxon:382"
     gene            complement(147. .1382)
                     /gene="SMc04259"
     CDS             complement(147. .1382)
                     /gene="SMc04259"
                     /function="cell processes; transport of small molecules"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PUTATIVE PERIPLASMIC BINDING ABC TRANSPORTER
                     SIGNAL PEPTIDE PROTEIN"
                     /protein_id="CAC46508.1"
                     /db_xref="GI:15074951"
                     /translation="MNLRFLAAALGATAALPFGAASATDLEVTHWWTSGGEAAAVAEL
                     AKAFDATGNKWVDGAIAGSGGTARPIMISRITGGDPMAATQFNHGRQAEELVQAGLMR
                     DLTDIATKENWKEIVKPSSLLDSCTIEGKIYCAPVNIHSWQWLWLSNAAFKQAGVEVP
                     KNWDEFVAAAPALEKAGIVPLAVGGQPWQANGAFDVLMVAIAGKENFEKVFAQKDEEV
                     AAGPEIAKVFKAADDARRMSKGTNVQDWNQATNMVITGKAGGQIMGDWAQGEFQLAGQ
                     KAGVDYTCLPGLGVNEVISTGGDAFYFPLLEDEEKSKAQEVLASTLLKPETQVAFNLK
                     KGSLPVRGDVDLAAANDCMKKGLDILAKGNVIQGTDQLLSADSQKQKEDLFSEFFANH
                     SMTPEDAQKRFADIIAAAD"
     gene            1858. .2928
                     /gene="SMc04260"
     CDS             1858. .2928
                     /gene="SMc04260"
                     /function="miscellaneous; not classified regulator"
                     /note="Product confidence : putative
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
                     /protein_id="CAC46509.1"
                     /db_xref="GI:15074952"
                     /translation="MTGMDGKTKNRATTAAPVEGGRPTLKTIAFMTGLGVTTVSRALK
                     DAPDIGAETKERVRLVAKQIGYQPNRAGVRLRTGKTNVISLVLTLEEEIMGITSPMVI
                     GITEILAGTQYHLVVTPYSSAKDPLGPIRYILDTGAADGVIISRTEPNDPRVTLLTER
                     RLPFATHGRTEMGLIHPYHDFDNERFAYEAVRKLVDRGRRRLVLLEPPPNLTFHSHMR
                     TGFERGVRDFGAEAVSFHQVNIDHSLVAIRDAFEKLMRSPDAPDGVVSGSGSGAIALI
                     AGIEAAGKKVGEDAEMVSKVPSDFLRWLRPEVMTMYEDIRIAGRELAKAVIGHIEGRP
                     PETLQSLSQPEFQPPMARSAVS"
     repeat_region   2954. .3023
                     /note="Sm-4 OR SMc04506
                     REPEAT SM-4
                     predicted by Homology"
                     /evidence=not_experimental
     gene            complement(3036. .3302)
                     /gene="SMc04261"
     CDS             complement(3036. .3302)
                     /gene="SMc04261"
                     /function="miscellaneous; unknown"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by FrameD"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="HYPOTHETICAL PROTEIN"
                     /protein_id="CAC46510.1"
                     /db_xref="GI:15074953"
                     /translation="MKFTVIEGGKAAEVTAAGSVTIRDEAERRIKAAGYEQWRVRSLA
                     TGAPIPGPIQYLKMQIDFVAGKLEQLNPIPSDFTDDKYWPAAGI"
     gene            complement(3407. .4837)
                     /gene="gnd OR SMc04262"
     CDS             complement(3407. .4837)
                     /gene="gnd OR SMc04262"
                     /EC_number="1.1.1.44"
                     /function="small molecule metabolism; energy metabolism,
                     carbon; oxidative branch, pentose pathway"
                     /note="Product confidence : probable
```

```
                Gene name confidence : putative
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PROBABLE 6-PHOSPHOGLUCONATE DEHYDROGENASE
                (DECARBOXYLATING) PROTEIN"
                /protein_id="CAC46511.1"
                /db_xref="GI:15074954"
                /translation="MSQAEIGLIGLGVMGSNLALNIAEKGNRIAVFNRTVDATRKFYA
                EAGALKDQIVPCETIEEFVAAIRPPRPIIIMIKAGDPVDQQMEALKPHLAKGDIMIDA
                GNANFRDTMRRFDALKDSGLTFIGMGVSGGEEGARHGPSIMVGGTEESYRRVEKVLTS
                IAAKYDSDPCVAWLGENGAGHFVKTIHNGIEYADMQMIAEIYGILRDGLKMTAQEIGE
                VFGAWNKGRLNSYLIEITEKVLKAADPLTGKPIVDMILDKAGQKGTGKWSVIEAQNMG
                VPATAIEAAVAARSISSAKEEREAAEKVLGLPPVGEIKVADRDGFIKDLENALLAAKI
                GAYAQGFAVMAAASKEFGWNLPMPTIAKIWRAGCIIRSQFLDEITTAFTKAPDAANLI
                VTPAFAAMVKESDGALRRVVSTAVLGGLPVPALASALGYFDSYRRGRGTANVIQAQRD
                FFGAHGFDRVDGADSHHGPWGSGLNA"
gene            5273. .6691
                /gene="SMc04263"
CDS             5273. .6691
                /gene="SMc04263"
                /function="cell processes; transport of small molecules;
                amino acids, amines, peptides"
                /note="Product confidence : putative
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PUTATIVE AMINO ACID CARRIER TRANSMEMBRANE
                PROTEIN"
                /protein_id="CAC46512.1"
                /db_xref="GI:15074955"
                /translation="MDTIIGFLNTIFWGYVLIYGLLAVGLYFTVRLGFPQIVHFSEMF
                RVLSRGGSKDAAGISPFQALMVSLASRVGTGNLAGVAVALYLGGPGATFWMWIVAFVG
                MATAYAESALAQLYKIRNEDGQYRGGPAFYIARGLNAPWAATIFSACLILSFGLVFNA
                VQANSIADAVQGAFGVPKLAVGVAVALLSGVVIFGGIRQIARVAEIVVPFMAAAYLLT
                AVYVLIANASAVPGVLWTIISSAFGFQEAAGGITGGIAAAMLNGVKRGLFSNEAGMGS
                APNIAAVATPVPHHPSSQGFVQSLGVFIDTILICTATSVMILLSGTLEPGSGVTGTQL
                TQAAMSTHIGSAGTYFIAIAIFFFAFTSIIGNYSYAENALTYLGGGNRLGLMIMRCAT
                LAMVVWGAYESITTVFDAADASMGLMATINLVAILLLSGTIAKLTKDYFEQRKAGAVP
                VFHAADYPELQGKIDGEIWSRD"
gene            6773. .7663
                /gene="SMc04264"
CDS             6773. .7663
                /gene="SMc04264"
                /function="miscellaneous; not classified regulator"
                /note="Product confidence : putative
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
                /protein_id="CAC46513.1"
                /db_xref="GI:15074956"
                /translation="MGTSCPAYDKDARAPERLQQEARLRLGPPALAQTSGIYREQSPP
                AALAFHFKCLWLHRMPDGAPSSIAVVPDGCSDIIWSSKGLAIVGPDRIAAFPELPAGE
                TIIGARFRIGAAASWLRTPLREMTGQTVPLHLVWGPATNEIEARIRDAASTPERLGVL
                AAALLGRLSAAEPPPPDIAACVAHLQVGRSTVDDAIRTLVREIGTSERTLRRHCHEHL
                GYGAKTLGRILRLQRFLAACRSRPGDTLANLAVDAGYADQAHLAREARGLTSFTPREI
                QRQLAARFGIGRVENFRTTT"
gene            7796. .8239
                /gene="SMc04265"
CDS             7796. .8239
                /gene="SMc04265"
                /function="miscellaneous; hypothetical/global homology"
                /note="Product confidence : hypothetical
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"

  Query Match             72.8%;  Score 18.2;  DB 1;  Length 294800;
  Best Local Similarity   87.0%;  Pred. No. 3.1e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||  || ||
Db 165819 GTCAGGTGCAGGTCAGGGCGCTG 165797


RESULT  36
AE004192
LOCUS       AE004192    12891 bp    DNA             BCT       31-JUL-2000
DEFINITION  Vibrio cholerae chromosome I, section 100 of 251 of the complete
            chromosome.
ACCESSION   AE004192 AE003852
VERSION     AE004192.1  GI:9655581
KEYWORDS    .
SOURCE      Vibrio cholerae.
  ORGANISM  Vibrio cholerae
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 12891)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C. and White,O.
  TITLE     DNA sequence of both chromosomes of the cholera pathogen Vibrio
            cholerae
  JOURNAL   Nature 406 (6795), 477-483 (2000)
  MEDLINE   20406833
REFERENCE   2  (bases 1 to 12891)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
            Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .12891
                     /organism="Vibrio cholerae"
                     /strain="N16961"
                     /serotype="O1"
                     /db_xref="taxon:666"
                     /chromosome="I"
                     /note="biotype: El Tor"
     gene            complement(84. .1370)
                     /gene="VC1111"
     CDS             complement(84. .1370)
                     /gene="VC1111"
                     /note="similar to GB:J04423 SP:P12995 PID:457106
                     PID:490217 GB:U00096; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="adenosylmethionine-8-amino-7-oxononanoate
                     aminotransferase"
                     /protein_id="AAF94270.1"
                     /db_xref="GI:9655582"
                     /translation="MMMDLAFDRCHIWHPYTSTLTPLTCYPVSRADGVMITLEDGRSL
                     VDGMSSWWAAIHGYNHPKLNQAAHQQIEQMSHIMFGGLTHQPAIELCQKLLKLAPNNL
                     QQVFLADSGSVAVEVSLKMALQYWHSKGQPRAKFLTLRHGYHGDTFAAMSVTDPDNSM
```

```
                HSLYKGFLPEHIFANSPEGGFFDAWDERDIADFRHKLTEHHHQIAAVILEPIVQGAGG
                MRIYHPEFLRQVRALCDEFGVLLILDEIATGFGRTGKLFACEHAGIQPDILCVGKALT
                GGYMTLSATLTTQQVADTVCAGEAGCFMHGPTFMGNPLACAVASASLSLIEQGDWQQQ
                VANIEAFFAEQLPKLNDSPRVKQTRWLGAIGVVETHLPVNMEVIQALFVEHGVWIRPF
                GRLIYLMPPYISQPEHLQKLVDAMAAALNQPDCFKQ"
gene            1497. .2549
                /gene="VC1112"
CDS             1497. .2549
                /gene="VC1112"
                /note="similar to GB:J04423 SP:P12996 PID:145425
                PID:490219 GB:U00096; identified by sequence similarity;
                putative"
                /codon_start=1
                /transl_table=11
                /product="biotin synthase"
                /protein_id="AAF94271.1"
                /db_xref="GI:9655583"
                /translation="MEVRHNWTVAEVKALLDKPFMDLLFEAQQVHRLHHPHNHVQVST
                LLSIKTGACPEDCKYCPQSAHYRTDVDKERLMEVERVLDAAQKAKNSGSTRFCMGAAW
                KNPKERDMPLLKEMIRGVKDMGLETCMTLGMLTPDQAQQLAQAGLDYYNHNLDTSPEF
                YGNIITTRTYQDRLDTLSHVRDAGMKICSGGIIGMGESTNDRAGLLVELANLPTHPES
                VPINMLVKVKGTPLEQVDDVEPFDFVRLIAVARIMMPKSAVRLSAGREKMNEQMQALC
                FMAGANSIFYGCKLLTTPNPAEDSDMLLFKKLGINREQVAQKPDEITENELLDRVVER
                VAARPTASDLFYDAAL"
gene            2533. .3687
                /gene="VC1113"
CDS             2533. .3687
                /gene="VC1113"
                /note="similar to GB:J04423 SP:P12998 PID:145426
                PID:490225 GB:U00096; identified by sequence similarity;
                putative"
                /codon_start=1
                /transl_table=11
                /product="8-amino-7-oxononanoate synthase"
                /protein_id="AAF94272.1"
                /db_xref="GI:9655584"
                /translation="MTLRFKSRIEQALNEREHSRLTRRLQLLDRSAQGELIQQDRTYL
                NFSSNDYLGLANDAELVQAWQTGLARFGAGSGASPMVTGFSSAHAELEHELCHWLGYE
                RAVLFSSGFSANQALLFTLLEKEDLLLQDKLNHASLMEAGMLSPATMKRFKHNDTEHL
                RQLLHEQSNSLVVTEGVFSMDGDCAPLAQIATLVKQRAWLMVDDAHGIGVLGEDGAGS
                CQAAGIHPEILVVTFGKAFGLAGAAVLCDAHVGDYLTQFARHHVYSTAMPPAQAHALT
                HAARMIQSQSWRREQLAELLACFDEQCRNIPGFVATQTPIKPWLLGSSDSALSASHEL
                KQQGIWVSAIRPPTVPVGSARLRITITAAHTQAQIRRLSEQLKHVMEKDT"
gene            3549. .4487
                /gene="VC1114"
CDS             3549. .4487
                /gene="VC1114"
                /note="similar to GB:J04423 SP:P12999 PID:145427
                PID:490221 GB:U00096; identified by sequence similarity;
                putative"
                /codon_start=1
                /transl_table=11
                /product="biotin synthesis protein BioC"
                /protein_id="AAF94273.1"
                /db_xref="GI:9655585"
                /translation="MGECHSTTHRACRQCAFTYHDNGGAYPSADTPLIGAVKTCDGEG
                HMSMTATELCVELKDKSAIAQAFGKAAAHYDQHAAFQRDVGLRLLQKMPSCLKGLRVL
                DLGCGTGYFSALLRERGAQVVCADISHAMLEQAKQRCGDEGMSYQLADAEQLPFASAC
                FDMVFSSLALQWCEDLSLPLSEIRRVLKPHGQAFLSTLLDGSLFELEQAWRSVDHHRH
                INQFISINQVKIALAQAGCSQHHLDLAAITVWYEAAFMLMRDLKGIGANHVSGRSTGL
                ISRRTLAKVELAYQSFRNQQGLVPATYQVCLGVIHR"
gene            4463. .5179
                /gene="VC1115"
CDS             4463. .5179
                /gene="VC1115"
                /note="similar to GB:S68059 SP:P13000 PID:490223
                PID:145429 PID:460901; identified by sequence similarity;
                putative"
                /codon_start=1
                /transl_table=11
                /product="dethiobiotin synthetase"
                /protein_id="AAF94274.1"
                /db_xref="GI:9655586"
                /translation="MFRGYSPMRNAIFIAGTDTDVGKTVASKAILQALATHNIATIGY
                KPVAAGSDKTEFGYRNSDALHLMKAATVDMPYEDVNPYALVLPTSPHIAAKHENVTID
                YALLSNKLSKHKQNAELVVVEGAGGWRVPTSDSDCLSTWVKQERLPVILVVGIKLGCL
                SHAILTAEAIRADGLELVGWIANRINPGTEHYAEIIEHLEGRLGTPKLGEIPYMPKAK
                RQELGKFIQLDHLLEPDTVA"
gene            complement(5261. .5470)
                /gene="VC1116"
CDS             complement(5261. .5470)
                /gene="VC1116"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF94275.1"
                /db_xref="GI:9655587"
                /translation="MSRRNLTTNKGKHMAEAHSYQMAIDLLCCHLGISEQEARTQLGL
                EAKEQTLQQQIVDTQHSLMGLSAEK"
gene            5697. .6560
                /gene="VC1117"
CDS             5697. .6560
                /gene="VC1117"
                /note="similar to GB:M58470 SP:P23894 PID:146412 GB:U00096
                PID:1736470; identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="heat shock protein HtpX"
                /protein_id="AAF94276.1"
                /db_xref="GI:9655588"
                /translation="MKRILLFLATNLAVVLVLSVVLNIVYAVTGMQPGSLSGLLVMAA
                VFGFGGAFISLLMSKSMALRSVGGVVIDTPRNEMEHWLLETVRRQANQAGIGMPTVAI
                YDAPDMNAFATGAKRDDSLVAVSTGLLHNMTRDEAEAVLAHEVSHIANGDMVTMTLMQ
                GVVNTFVIFLSRFIANIVASRDSEEGEGSNMMVYFGVSMVLELVFGFLASFITMWYSR
                HREFHADAGAAQLVGKHKMIAALERLKMGQESHLEGSMMAFGITGKRSLSELMMTHPP
                LEKRIAALRNM"
gene            6770. .7198
                /gene="VC1118"
CDS             6770. .7198
                /gene="VC1118"
                /note="similar to PID:1653608; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="transcriptional regulator, putative"
                /protein_id="AAF94277.1"
                /db_xref="GI:9655589"
                /translation="MSIMDVQHVAQFYQQLDKSQLHRLIEIYHPDVVFEDAAHRIEGF
                DALYQYFLNLYQNVHTCTFTIHEQYAVNEGAFLVWTMHLRHPKLAKGEQVDVKGVSHL
                HFAEGKVTYHRDYFDMGEMLYEQLPVLGQVIRAIKRRLGQ"
gene            7195. .7920
                /gene="VC1119"
CDS             7195. .7920
                /gene="VC1119"
                /note="similar to SP:P25970; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="oxidoreductase, short-chain
                dehydrogenase/reductase family"
                /protein_id="AAF94278.1"
                /db_xref="GI:9655590"
                /translation="MTAVFITGATSGIGKQLALDYAKQGWQVIACGRNQSVLDSLHTQ
                YANIFPLAFDVTDHPGTKAALAQLPCQPELWILNAGDCEYIDDGKMDVTLMARVFNIN
                VLGVAAVIEGIQPHLSCGHRVAIVGSIASELALPRAEAYGASKAAVAYLARTLQLDWR
                PLGIEVTTIFPGFVATPLTDRNTFAMPMIITVERAAQEIKAGLARGVSQLYFPKRFTW
                LIRLLGALPYAWQGRLVRRLLKA"
gene            7931. .9211
                /gene="VC1120"
CDS             7931. .9211
                /gene="VC1120"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
```

```
  Query Match             71.2%;  Score 17.8;  DB 1;  Length 12891;
  Best Local Similarity   90.5%;  Pred. No. 7.3e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagcacgttg 25
           || |||||||||||||| ||
Db    1611 CATGTGCAGGTCAGCACGCTG 1631


RESULT  37
AF145866/c
LOCUS       AF145866      694 bp    DNA             BCT       10-JUN-1999
DEFINITION  Metal-contaminated soil clone K20-78 16S ribosomal RNA gene,
            partial sequence.
ACCESSION   AF145866
VERSION     AF145866.1  GI:5031368
KEYWORDS    .
SOURCE      metal-contaminated soil clone K20-78.
  ORGANISM  metal-contaminated soil clone K20-78
            Bacteria; Green non-sulfur bacteria; environmental samples.
REFERENCE   1  (bases 1 to 694)
  AUTHORS   Marsh,T.L., Kurzhals,R.L., Dojka,M.A., Icopini,G., Long,D.T. and
            Pace,N.R.
  TITLE     Molecular diversity of a soil bacterial community from a metal
            contaminated super fund site
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 694)
  AUTHORS   Marsh,T.L., Kurzhals,R.L., Dojka,M.A., Icopini,G., Long,D.T. and
            Pace,N.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-1999) Plant and Microbial Biology, University of
            California, Berkeley, 111 Koshland Hall, Berkeley, CA 94720, USA
FEATURES             Location/Qualifiers
     source          1. .694
                     /organism="metal-contaminated soil clone K20-78"
                     /db_xref="taxon:95834"
                     /clone="K20-78"
     rRNA            <1. .>694
                     /product="16S ribosomal RNA"
BASE COUNT      179 a    150 c    237 g    128 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 1;  Length 694;
  Best Local Similarity   83.3%;  Pred. No. 1.4e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           |||||| |||||  |||||||| |
Db     679 CGTCAGATGCAGCCCAGCACGTCG 656


RESULT  38
AF269004/c
LOCUS       AF269004     1412 bp    DNA             BCT       16-FEB-2001
DEFINITION  Uncultured bacterium SBR1108 16S ribosomal RNA gene, partial
            sequence.
ACCESSION   AF269004 X84534
VERSION     AF269004.1  GI:12751010
KEYWORDS    .
SOURCE      uncultured bacterium SBR1108.
  ORGANISM  uncultured bacterium SBR1108
            Bacteria; Green non-sulfur bacteria; environmental samples.
REFERENCE   1  (bases 1 to 1412)
  AUTHORS   Bond,P.L., Hugenholtz,P., Keller,J. and Blackall,L.L.
  TITLE     Bacterial community structures of phosphate-removing and
            non-phosphate-removing activated sludges from sequencing batch
            reactors
  JOURNAL   Appl. Environ. Microbiol. 61 (5), 1910-1916 (1995)
  MEDLINE   95374005
   PUBMED   7544094
REFERENCE   2  (bases 1 to 1412)
  AUTHORS   Hugenholtz,P., Tyson,G.W., Webb,R.I., Wagner,A.M. and Blackall,L.L.
  TITLE     Investigation of candidate division TM7, a recently recognized
            major lineage of the domain bacteria with no known pure-culture
            representatives
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 411-419 (2001)
  MEDLINE   20575237
   PUBMED   11133473
REFERENCE   3  (bases 1 to 1412)
  AUTHORS   Bond,P.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-1995) Department of Chemical Engineering,
            University of Queensland, Brisbane, Queensland 4072, Australia
REFERENCE   4  (bases 1 to 1412)
  AUTHORS   Hugenholtz,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2000) Microbiology and Parasitology, The
            University of Queensland, Molecular Biosciences Building 76, Cooper
            Rd., Brisbane, Queensland 4072, Australia
COMMENT     On Feb 16, 2001 this sequence version replaced gi:871655.
FEATURES             Location/Qualifiers
     source          1. .1412
                     /organism="uncultured bacterium SBR1108"
                     /db_xref="taxon:142194"
                     /clone="SBR1108"
                     /note="clone obtained from a sequencing batch reactor"
     rRNA            <1. .>1412
                     /product="16S ribosomal RNA"
BASE COUNT      359 a    332 c    467 g    254 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 1;  Length 1412;
  Best Local Similarity   83.3%;  Pred. No. 1.2e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           |||||||||||  ||| |||| |
Db     689 CGTCAGGTGCAGACCAGAACGTCG 666


RESULT  39
HUMHXBA/c
LOCUS       HUMHXBA      4877 bp    mRNA            PRI       08-NOV-1994
DEFINITION  Human hexabrachion mRNA, partial cds.
ACCESSION   M24630
VERSION     M24630.1  GI:514363
KEYWORDS    cytoactin; hexabrachion; tenascin.
SOURCE      Homo sapiens (tissue library: ATCC HTB-17) glioblastoma cDNA to
            mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4877)
  AUTHORS   Gulcher,J.R., Nies,D.E., Marton,L.S. and Stefansson,K.
  TITLE     An alternatively spliced region of the human hexabrachion contains
            a repeat of potential N-glycosylation sites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 86 (5), 1588-1592 (1989)
  MEDLINE   89160821
COMMENT     On Jul 22, 1994 this sequence version replaced gi:341208.
FEATURES             Location/Qualifiers
     source          1. .4877
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="U-373MG"
                     /tissue_type="glioblastoma"
                     /tissue_lib="ATCC HTB-17"
                     /map="9q32-q34"
     gene            1. .4877
                     /gene="HXB"
     CDS             <1. .>4877
                     /gene="HXB"
                     /codon_start=1
                     /db_xref="GDB:G00-120-073"
                     /product="hexabrachion"
```

```
                     /protein_id="AAA52703.1"
                     /db_xref="GI:553348"
                     /translation="GEDCSQLRCPNDCHSRGRCVEGKCVCEQGFKGYDCSDMSCPNDC
                     HQHGRCVNGMCVCDDGYTGEDCRDRQCPRDCSNRGLCVDGQCVCEDGFTGPDCAELSC
                     PNDCHGRGRCVNGQCVCHEGFMGKDCKEQRCPSDCHGQGRCVDGQCICHEGFTGLDCG
                     QHSCPSDCNNLGQCVSGRCICNEGYSGEDCSEVSPPKDLVVTEVTEETVNLAWDNEMR
                     VTEYLVVYTPTHEGGLEMQFRVPGDQTSTIIQELEPGVEYFIRVFAILENKKSIPVSA
                     RVATYLPAPEGLKFKSIKETSVEVEWDPLDIAFETWEIIFRNMNKEDEGEITKSLRRP
                     ETSYRQTGLAPGQEYEISLHIVKNNTRGPGLKRVTTTRLDAPSQIEVKDVTDTTALIT
                     WFKPLAEIDGIELTYGIKDVPGDRTTIDLTEDENQYSIGNLKPDTEYEVSLISRRGDM
                     SSNPAKETFTTGLDAPRNLRRVSQTDNSITLEWRNGKAAIDSYRIKYAPISGGDHAEV
                     DVPKSQQATTKTTLTGLRPGTEYGIGVSAVKEDKESNPATINAATELDTPKDLQVSET
                     AETSLTLLWKTPLAKFDRYRLNYSLPTGQWVGVQLPRNTTSYVLRGLEPGQEYNVLLT
                     AEKGRHKSKPAHVKASTEQAPELENLTVTEVGWDGLRLNWTAADQAYEHFIIQVQEAN
                     KVEAARNLTVPGSLRAVDIPGLKAATPYTVSIYGVIQGYRTPVLSAEASTGETPNLGE
                     VVVAEVGWDALKLNWTAPEGAYEYFFIQVQEADTVEAAQNLTVPGGLRSTDLPGLKAA
                     THYTITIRGVTQDFSTTPLSVEVLTEEVPDMGNLTVTEVSWDALRLNWTTPDGTYDQF
                     TIQVQEADQVEEAHNLTVPGSLRSMEIPGLRAGTPYTVTLHGEVRGHSTRPLAVEVVT
                     EDLPQLGDLAVSEVGWDGLRLNWTAADNAYEHFVIQVQEVNKVEAAQNLTLPGSLRAV
                     DIPGLEAATPYRVSIYGVIRGYRTPVLSAEASTAKEPEIGNLNVSDITPESFNLSWMA
                     TDGIFETFTIEIIDSNRLLETVEYNISGAERTAHISGLPPSTDFIVYLSGLAPSIRTK
                     TISATATTEALPLLENLTISDINPYGFTVSWMASENAFDSFLVTVVDSGKLLDPQEFT
                     LSGTQRKLELRGLITGIGYEVMVLWLHPRASNKPLRAEIVTEAEPEVDNLLVSDATPD
                     GFRLSWTADEGVFDNFVLKIRDTKKQSEPLEITLLAPERTRDLTGLREATEYEIELYG
                     ISKGRRSQTVSAIATTAMGSPKEVIFSDITENSATVSWRAPTAQVESFRITYVPITGG
                     TPSMVTVDGTKTQTRLVKLIPGVEYLVSIIAMKGFEESEPVSGSFTTALDGPSGLVTA
                     NITDSEALARWQPAIATVDSYVISYTGEKVPEITRTVSGNTVEYALTDLEPATEYTLR
                     IFAEKGPQKSSTITAKFTTDLDSPRDLTATEVQSETALLTWRPPRASVTGYLLVYESV
                     DGTVKEVIVGPDTTSYSLADLSPSTHYTAKIQALNGPLRSNMIQTIFTTIGLLYPFPK
                     DCSQAMLNGDTTSGLYTIYLNGDKAQALEVFCDMTSDGGGWIVFLRRKNGRENFYQNW
                     KAYAAGFGDRREEFL"
BASE COUNT     1239 a   1345 c   1311 g    982 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 9;  Length 4877;
  Best Local Similarity   83.3%;  Pred. No. 1e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           | ||||||||||| || |||| |
Db     859 CTTCAGGTGCAGGTAAGTACGTGG 836


RESULT  40
HSTNSCN/c
LOCUS       HSTNSCN      7286 bp    mRNA            PRI       04-JUN-1998
DEFINITION  Human mRNA for tenascin.
ACCESSION   X56160
VERSION     X56160.1  GI:37226
KEYWORDS    extracellular matrix protein; tenascin.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7280)
  AUTHORS   Zardi,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-1990) L. Zardi, ISTITUTO NAZIONALE PER LA RICERCA
            SUL, CANCRO VIALE BENETTO XV. 10, 16132 GENOVA, ITALY
REFERENCE   2  (bases 1 to 7280)
  AUTHORS   Siri,A., Carnemolla,B., Saginati,M., Leprini,A., Casari,G.,
            Baralle,F. and Zardi,L.
  TITLE     Human tenascin: primary structure, pre-mRNA splicing patterns and
            localization of the epitopes recognized by two monoclonal
            antibodies
  JOURNAL   Nucleic Acids Res. 19 (3), 525-531 (1991)
  MEDLINE   91187670
COMMENT     This sequence conflicts with M24630
            (Gulcher J.R. P.N.A.S. 86:1588-1592(1989).
FEATURES             Location/Qualifiers
     source          1. .7286
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9q33-q34"
                     /dev_stage="adult & fetal"
                     /tissue_type="melanoma, brain, breast"
                     /tissue_lib="human fetal brain library in lambda gt11,
                     human malanoma library"
     mRNA            7. .7286
     CDS             55. .6654
                     /codon_start=1
                     /product="tenascin"
                     /protein_id="CAA39628.1"
                     /db_xref="GI:37227"
                     /db_xref="SWISS-PROT:P24821"
                     /translation="MGAMTQLLAGVFLAFLALATEGGVLKKVIRHKRQSGVNATLPEE
                     NQPVVFNHVYNIKLPVGSQCSVDLESASGEKDLAPPSEPSESFQEHTVDGENQIVFTH
                     RINIPRRACGCAAAPDVKELLSRLEELENLVSSLREQCTAGAGCCLQPATGRLDTRPF
                     CSGRGNFSTEGCGCVCEPGWKGPNCSEPECPGNCHLRGRCIDGQCICDDGFTGEDCSQ
                     LACPSDCNDQGKCVNGVCICFEGYAADCSREICPVPCSEEHGTCVDGLCVCHDGFAGD
                     DCNKPLCLNNCYNRGRCVENECVCDEGFTGEDCSELICPNDCFDRGRCINGTCYCEEG
                     FTGEDCGKPTCPHACHTQGRCEEGQCVCDEGFAGVDCSEKRCPADCHNRGRCVDGRCE
                     CDDGFTGADCGELKCPNGCSGHGRCVNGQCVCDEGYTGEDCSQLRCPNDCHSRGRCVE
                     GKCVCEQGFKGYDCSDMSCPNDCHQHGRCVNGMCVCDDGYTGEDCRDRQCPRDCSNRG
                     LCVDGQCVCEDGFTGPDCAELSCPNDCHGQGRCVNGQCVCHEGFMGKDCKEQRCPSDC
                     HGQGRCVDGQCICHEGFTGLDCGQHSCPSDCNNLGQCVSGRCICNEGYSGEDCSEVSP
                     PKDLVVTEVTEETVNLAWDNEMRVTEYLVVYTPTHEGGLEMQFRVPGDQTSTIIRELE
                     PGVEYFIRVFAILENKKSIPVSARVATYLPAPEGLKFKSIKETSVEVEWDPLDIAFET
                     WEIIFRNMNKEDEGEITKSLRRPETSYRQTGLAPGQEYEISLHIVKNNTRGPGLKRVT
                     TTRLDAPSQIEVKDVTDTTALITWFKPLAEIDGIELTYGIKDVPGDRTTIDLTEDENQ
                     YSIGNLKPDTEYEVSLISRRGDMSSNPAKETFTTGLDAPRNLRRVSQTDNSITLEWRN
                     GKAAIDSYRIKYAPISGGDHAEVDVPKSQQATTKTTLTGLRPGTEYGIGVSAVKEDKE
                     SNPATINAATELDTPKDLQVSETAETSLTLLWKTPLAKFDRYRLNYSLPTGQWVGVQL
                     PRNTTSYVLRGLEPGQEYNVLLTAEKGRHKSKPARVKASTEQAPELENLTVTEVGWDG
                     LRLNWTAADQAYEHFIIQVQEANKVEAARNLTVPGSLRAVDIPGLKAATPYTVSIYGV
                     IQGYRTPVLSAEASTGETPNLGEVVVAEVGWDALKLNWTAPEGAYEYFFIQVQEADTV
                     EAAQNLTVPGGLRSTDLPGLKAATHYTITIRGVTQDFSTTPLSVEVLTEEVPDMGNLT
                     VTEVSWDALRLNWTTPDGTYDQFTIQVQEADQVEEAHNLTVPGSLRSMEIPGLRAGTP
                     YTVTLHGEVRGHSTRPLAVEVVTEDLPQLGDLAVSEVGWDGLRLNWTAADNAYEHFVI
                     QVQEVNKVEAAQNLTLPGSLRAVDIPGLEAATPYRVSIYGVIRGYRTPVLSAEASTAK
                     EPEIGNLNVSDITPESFNLSWMATDGIFETFTIEIIDSNRLLETVEYNISGAERTAHI
                     SGLPPSTDFIVYLSGLAPSIRTKTISATATTEALPLLENLTISDINPYGFTVSWMASE
                     NAFDSFLVTVVDSGKLLDPQEFTLSGTQRKLELRGLITGIGYEVMVSGFTQGHQTKPL
                     RAEIVTEAEPEVDNLLVSDATPDGFRLSWTADEGVFDNFVLKIRDTKKQSEPLEITLL
                     APERTRDITGLREATEYEIELYGISKGRRSQTVSAIATTAMGSPKEVIFSDITENSAT
                     VSWRAPTAQVESFRITYVPITGGTPSMVTVDGTKTQTRLVKLIPGVEYLVSIIAMKGF
                     EESEPVSGSFTTALDGPSGLVTANITDSEALARWQPAIATVDSYVISYTGEKVPEITR
                     TVSGNTVEYALTDLEPATEYTLRIFAEKGPQKSSTITAKFTTDLDSPRDLTATEVQSE
                     TALLTWRPPRASVTGYLLVYESVDGTVKEVIVGPDTTSYSLADLSPSTHYTAKIQALN
                     GPLRSNMIQTIFTTIGLLYPFPKDCSQAMLNGDTTSGLYTIYLNGDKAQALEVFCDMT
                     SDGGGWIVFLRRKNGRENFYQNWKAYAAGFGDRREEFWLGLDNLNKITAQGQYELRVD
                     LRDHGETAFAVYDKFSVGDAKTRYKLKVEGYSGTAGDSMAYHNGRSFSTFDKDTDSAI
                     TNCALSTRGFWYRNCHRVNLMGRYGDNNHSQGVNWFHWKGHEHSIQFAEMKLRPSNFR
                     NLEGRRKRA"
     repeat_region   574. .1911
                     /rpt_family="EGF-like"
                     /rpt_unit=603. .696
                     /rpt_type=FLANKING
                     /evidence=experimental
     repeat_region   1912. .5970
                     /rpt_family="EGF-like"
                     /rpt_unit=1906. .2175
                     /rpt_type=FLANKING
                     /evidence=experimental
BASE COUNT     1830 a   1926 c   1999 g   1531 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 9;  Length 7286;
  Best Local Similarity   83.3%;  Pred. No. 9.8e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           | ||||||||||| || |||| |
Db    2200 CTTCAGGTGCAGGTAAGTACGTGG 2177
```

```
RESULT  41
AX022893/c
LOCUS       AX022893     7346 bp    DNA             PAT       07-SEP-2000
DEFINITION  Sequence 1 from Patent WO9925819.
ACCESSION   AX022893
VERSION     AX022893.1  GI:10046384
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 7346)
  AUTHORS   Uhlmann,E., Weiser,C. and Peyman,A.
  TITLE     Antisense oligonucleotides against tenascin for treating vitiligo
  JOURNAL   Patent: WO 9925819-A 1 27-MAY-1999;
            UHLMANN EUGEN (DE); WEISER CAROLINE (DE); HOECHST MARION ROUSSEL DE
            GMBH (DE); PEYMAN ANUSCHIRWAN (DE)
FEATURES             Location/Qualifiers
     source          1. .7346
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     exon            1. .7346
BASE COUNT     1840 a   1945 c   2020 g   1541 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 6;  Length 7346;
  Best Local Similarity   83.3%;  Pred. No. 9.8e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           | ||||||||||| || |||| |
Db    2200 CTTCAGGTGCAGGTAAGTACGTGG 2177


RESULT  42
AX030481/c
LOCUS       AX030481     7346 bp    DNA             PAT       21-SEP-2000
DEFINITION  Sequence 1 from Patent DE19750702.
ACCESSION   AX030481
VERSION     AX030481.1  GI:10278038
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 7346)
  AUTHORS   Peyman,A.D., Uhlmann,E.D. and Weiser,C.D.
  TITLE     Antisense oligonucleotides that bind to sequences encoding human
            tenascin for treating depigmentation, cancer, inflammation and
            cardiovascular disease
  JOURNAL   Patent: DE 19750702-A 1 27-MAY-1999;
            HOECHST MARION ROUSSEL DE GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .7346
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     exon            1. .7346
BASE COUNT     1840 a   1945 c   2020 g   1541 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 6;  Length 7346;
  Best Local Similarity   83.3%;  Pred. No. 9.8e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           | ||||||||||| || |||| |
Db    2200 CTTCAGGTGCAGGTAAGTACGTGG 2177


RESULT  43
HUMHXB/c
LOCUS       HUMHXB       7390 bp    mRNA            PRI       14-FEB-1996
DEFINITION  Homo sapiens hexabrachion (HXB) mRNA, complete cds.
ACCESSION   M55618
VERSION     M55618.1  GI:184483
KEYWORDS    hexabrachion; tenascin.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7390)
  AUTHORS   Nies,D.E., Hemesath,T.J., Kim,J.H., Gulcher,J.R. and Stefansson,K.
  TITLE     The complete cDNA sequence of human hexabrachion (Tenascin). A
            multidomain protein containing unique epidermal growth factor
            repeats
  JOURNAL   J. Biol. Chem. 266 (5), 2818-2823 (1991)
  MEDLINE   91131572
REFERENCE   2  (bases 1 to 7390)
  AUTHORS   Gulcher,J.R., Nies,D.E., Alexakos,M.J., Ravikant,N.A.,
            Sturgill,M.E., Marton,L.S. and Stefansson,K.
  TITLE     Structure of the human hexabrachion (tenascin) gene
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88 (21), 9438-9442 (1991)
  MEDLINE   92052108
FEATURES             Location/Qualifiers
     source          1. .7390
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="9q32-q34"
     gene            1. .7390
                     /gene="HXB"
     exon            1. .592
                     /partial
                     /gene="HXB"
                     /note="G00-120-073"
                     /number=1
     CDS             136. .6747
                     /gene="HXB"
                     /codon_start=1
                     /db_xref="GDB:G00-120-073"
                     /product="hexabrachion"
                     /protein_id="AAA88083.1"
                     /db_xref="GI:184484"
                     /translation="MGAMTQLLAGVFLAFLALATEGGVLKKVIRHKRQSGVNATLPEE
                     NQPVVFNHVYNIKLPVGSQCSVDLESASGEKDLAPPSEPSESFQEHTVDGENQIVFTH
                     RINIPRRACGCAAAPDVKELLSRLEELENLVSSLREQCTAGAGCCLQPATGRLDTRPF
                     CSGRGNFSTEGCGCVCEPGWKGPNCSEPECPGNCHLRGRCIDGQCICDDGFTGEDCSQ
                     LACPSDCNDQGKCVNGVCICFEGYAGADCSREICPVPCSEEHGTCVDGLCVCHDGFAG
                     DDCNKPLCLNNCYNRGRCVENECVCDEGFTGEDCSELICPNDCFDRGRCINGTCYCEE
                     GFTGEDCGKPTCPHACHTQGRCEEGQCVCDEGFAGLDCSEKRCPADCHNRGRCVDGRC
                     ECDDGFTGADCGELKCPNGCSGHGRCVNGQCVCDEGYTGEDCSQLRCPNDCHSRGRCV
                     EGKCVCEQGFKGYDCSDMSCPNDCHQHGRCVNGMCVCDDGYTGEDCRDRQCPRDCSNR
                     GLCVDGQCVCEDGFTGPDCAELSCPNDCHGRGRCVNGQCVCHEGFMGKDCKEQRCPSD
                     CHGQGRCVDGQCICHEGFTGLDCGQHSCPSDCNNLGQCVSGRCICNEGYSGEDCSEVS
                     PPKDLVVTEVTEETVNLAWDNEMRVTEYLVVYTPTHEGGLEMQFRVPGDQTSTIIQEL
                     EPGVEYFIRVFAILENKKSIPVSARVATYLPAPEGLKFKSIKETSVEVEWDPLDIAFE
                     TWEIIFRNMNKEDEGEITKSLRRPETSYRQTGLAPGQEYEISLHIVKNNTRGPGLKRV
                     TTTRLDAPSQIEVKDVTDTTALITWFKPLAEIDGIELTYGIKDVPGDRTTIDLTEDEN
                     QYSIGNLKPDTEYEVSLISRRGDMSSNPAKETFTTGLDAPRNLRRVSQTDNSITLEWR
                     NGKAAIDSYRIKYAPISGGDHAEVDVPKSQQATTKTTLTGLRPGTEYGIGVSAVKEDK
                     ESNPATINAATELDTPKDLQVSETAETSLTLLWKTPLAKFDRYRLNYSLPTGQWVGVQ
                     LPRNTTSYVLRGLEPGQEYNVLLTAEKGRHKSKPAHVKASTEQAPELENLTVTEVGWD
                     GLRLNWTAADQAYEHFIIQVQEANKVEAARNLTVPGSLRAVDIPGLKAATPYTVSIYG
                     VIQGYRTPVLSAEASTGETPNLGEVVVAEVGWDALKLNWTAPEGAYEYFFIQVQEADT
                     VEAAQNLTVPGGLRSTDLPGLKAATHYTITIRGVTQDFSTTPLSVEVLTEEVPDMGNL
                     TVTEVSWDALRLNWTTPDGTYDQFTIQVQEADQVEEAHNLTVPGSLRSMEIPGLRAGT
                     PYTVTLHGEVRGHSTRPLAVEVVTEDLPQLGDLAVSEVGWDGLRLNWTAADNAYEHFV
                     IQVQEVNKVEAAQNLTLPGSLRAVDIPGLEAATPYRVSIYGVIRGYRTPVLSAEASTA
                     KEPEIGNLNVSDITPESFNLSWMATDGIFETFTIEIIDSNRLLETVEYNISGAERTAH
                     ISGLPPSTDFIVYLSGLAPSIRTKTISATATTEALPLLENLTISDINPYGFTVSWMAS
                     ENAFDSFLVTVVDSGKLLDPQEFTLSGTQRKLELRGLITGIGYEVMVLWLHPRASNKP
                     LRAEIVTEAEPEVDNLLVSDATPDGFRLSWTADEGVFDNFVLKIRDTKKQSEPLEITL
                     LAPERTRDLTGLREATEYEIELYGISKGRRSQTVSAIATTAMGSPKEVIFSDITENSA
                     TVSWRAPTAQVESFRITYVPITGGTPSMVTVDGTKTQTRLVKLIPGVEYLVSIIAMKG
                     FEESEPVSGSFTTALDGPSGLVTANITDSEALARWQPAIATVDSYVISYTGEKVPEIT
```

```
                RTVSGNTVEYALTDLEPATEYTLRIFAEKGPQKSSTITAKFTTDLDSPRDLTATEVQS
                ETALLTWRPPRASVTGYLLVYESVDGTVKEVIVGPDTTSYSLADLSPSTHYTAKIQAL
                NGPLRSNMIQTIFTTIGLLYPFPKDCSQAMLNGDTTSGLYTIYLNGDKAQALEVFCDM
                TSDGGGWIVFLRRKNGRENFYQNWKAYAAGFGDRREEFLHWLGLDNLNKITAQGQYEL
                RVDLRDHGETAFAVYDKFSVGDAKTRYKLKVEGYSGTAGDSMAYHNGRSFSTFDKDTD
                SAITNCALSYKGAFWYRNCHRVNLMGRYGDNNHSQGVNWFHWKGHEHSIQFAEMKLRP
                SNFRNLEGRRKRA"
     exon       593. .2002
                /gene="HXB"
                /note="G00-120-073"
                /number=2
     exon       2003. .2266
                /gene="HXB"
                /note="G00-120-073"
                /number=3
     exon       2267. .2382
                /gene="HXB"
                /note="G00-120-073"
                /number=4
     exon       2383. .2539
                /gene="HXB"
                /note="G00-120-073"
                /number=5
     exon       2540. .2809
                /gene="HXB"
                /note="G00-120-073"
                /number=6
     exon       2810. .2995
                /gene="HXB"
                /note="G00-120-073"
                /number=7
     exon       2996. .3085
                /gene="HXB"
                /note="G00-120-073"
                /number=8
     exon       3086. .3349
                /gene="HXB"
                /note="G00-120-073"
                /number=9
     exon       3350. .3622
                /gene="HXB"
                /note="G00-120-073"
                /number=10
     exon       3623. .3895
                /gene="HXB"
                /note="G00-120-073"
                /number=11
     exon       3896. .4168
                /gene="HXB"
                /note="G00-120-073"
                /number=12
     exon       4169. .4441
                /gene="HXB"
                /note="G00-120-073"
                /number=13
     exon       4442. .4714
                /gene="HXB"
                /note="G00-120-073"
                /number=14
     exon       4715. .4987
                /gene="HXB"
                /note="G00-120-073"
                /number=15
     exon       4988. .5260
                /gene="HXB"
                /note="G00-120-073"
                /number=16
     exon       5261. .5383
                /gene="HXB"
                /note="G00-120-073"
                /number=17
     exon       5384. .5527
                /gene="HXB"
                /note="G00-120-073"
                /number=18
     exon       5528. .5647
                /gene="HXB"
                /note="G00-120-073"
                /number=19
     exon       5648. .5791
                /gene="HXB"
                /note="G00-120-073"
                /number=20
     exon       5792. .5922
                /gene="HXB"
                /note="G00-120-073"
                /number=21
     exon       5923. .6055
                /gene="HXB"
                /note="G00-120-073"
                /number=22
     exon       6056. .6207
                /gene="HXB"
                /note="G00-120-073"
                /number=23
     exon       6208. .6310
                /gene="HXB"
                /note="G00-120-073"
                /number=24
     exon       6311. .6469
                /gene="HXB"
                /note="G00-120-073"
                /number=25
     exon       6470. .6636
                /gene="HXB"
                /note="G00-120-073"
                /number=26
     exon       6637. .7390
                /partial
                /gene="HXB"
                /note="G00-120-073"
                /number=27
BASE COUNT     1858 a   1954 c   2013 g   1565 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 9;  Length 7390;
  Best Local Similarity   83.3%;  Pred. No. 9.8e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacgttg 25
          | ||||||||||||  ||  |||| |
Db   2284 CTTCAGGTGCAGGTAAGTACGTGG 2261


RESULT  44
AX045676/c
LOCUS       AX045676     7560 bp    DNA             PAT       24-NOV-2000
DEFINITION  Sequence 3 from Patent WO0066628.
ACCESSION   AX045676
VERSION     AX045676.1  GI:11344051
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7560)
  AUTHORS   Geller,H.M. and Meiners,S.
  TITLE     Neurite outgrowth and guidance by tenascin-c
  JOURNAL   Patent: WO 0066628-A 3 09-NOV-2000;
            University of Medicine and Dentistry of New Jersey (US)
FEATURES             Location/Qualifiers
     source          1. .7560
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     1894 a   2019 c   2062 g   1585 t
```

```
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 6;  Length 7560;
  Best Local Similarity   83.3%;  Pred. No. 9.7e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           | ||||||||||| || |||| |
Db    2462 CTTCAGGTGCAGGTAAGTACGTGG 2439


RESULT  45
HSTENAS3/c
LOCUS       HSTENAS3     7560 bp    mRNA            PRI       02-MAY-1995
DEFINITION  H.sapiens mRNA for tenascin-C, 7560bp.
ACCESSION   X78565
VERSION     X78565.1  GI:556844
KEYWORDS    tenascin-C; wnascin.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7560)
  AUTHORS   Gherzi,R., Carnemolla,B., Siri,A., Ponassi,M., Balza,E. and
            Zardi,L.
  TITLE     Human tenascin gene. Structure of the 5'-region, identification,
            and characterization of the transcription regulatory sequences
  JOURNAL   J. Biol. Chem. 270 (7), 3429-3434 (1995)
  MEDLINE   95155442
REFERENCE   2  (bases 1 to 7560)
  AUTHORS   Zardi,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-1994) Luciano Zardi, Cell Biology Laboratory,
            Istituto Nazionale per la, Ricerca sul Cancro, Viale Benedetto XV,
            10, Genova, 16132, Italy
FEATURES             Location/Qualifiers
     source          1. .7560
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="Male"
                     /dev_stage="Adult, Fetal"
                     /tissue_type="melanoma, fetal brain, breast,  genomic from
                     placenta"
     CDS             314. .6919
                     /codon_start=1
                     /product="human tenascin-C"
                     /protein_id="CAA55309.1"
                     /db_xref="GI:556845"
                     /db_xref="SWISS-PROT:P24821"
                     /translation="MGAMTQLLAGVFLAFLALATEGGVLKKVIRHKRQSGVNATLPEE
                     NQPVVFNHVYNIKLPVGSQCSVDLESASGEKDLAPPSEPSESFQEHTVDGENQIVFTH
                     RINIPRRACGCAAAPDVKELLSRLEELENLVSSLREQCTAGAGCCLQPATGRLDTRPF
                     CSGRGNFSTEGCGCVCEPGWKGPNCSEPECPGNCHLRGRCIDGQCICDDGFTGEDCSQ
                     LACPSDCNDQGKCVNGVCICFEGYAGADCSREICPVPCSEEHGTCVDGLCVCHDGFAG
                     DDCNKPLCLNNCYNRGRCVENECVCDEGFTGEDCSELICPNDCFDRGRCINGTCYCEE
                     GFTGEDCGKPTCPHACHTQGRCEEGQCVCDEGFAGLDCSEKRCPADCHNRGRCVDGRC
                     ECDDGFTGADCGELKCPNGCSGHGRCVNGQCVCDEGYTGEDCSQLRCPNDCHSRGRCV
                     EGKCVCEQGFKGYDCSDMSCPNDCHQHGRCVNGMCVCDDGYTGEDCRDRQCPRDCSNR
                     GLCVDGQCVCEDGFTGPDCAELSCPNDCHGQGRCVNGQCVCHEGFMGKDCKEQRCPSD
                     CHGQGRCVDGQCICHEGFTGLDCGQHSCPSDCNNLGQCVSGRCICNEGYSGEDCSEVS
                     PPKDLVVTEVTEETVNLAWDNEMRVTEYLVVYTPTHEGGLEMQFRVPGDQTSTIIQEL
                     EPGVEYFIRVFAILENKKSIPVSARVATYLPAPEGLKFKSIKETSVEVEWDPLDIAFE
                     TWEIIFRNMNKEDEGEITKSLRRPETSYRQTGLAPGQEYEISLHIVKNNTRGPGLKRV
                     TTTRLDAPSQIEVKDVTDTTALITWFKPLAEIDGIELTYGIKDVPGDRTTIDLTEDEN
                     QYSIGNLKPDTEYEVSLISRRGDMSSNPAKETFTTGLDAPRNLRRVSQTDNSITLEWR
                     NGKAAIDSYRIKYAPISGGDHAEVDVPKSQQATTKTTLTGLRPGTEYGIGVSAVKEDK
                     ESNPATINAATELDTPKDLQVSETAETSLTLLWKTPLAKFDRYRLNYSLPTGQWVGVQ
                     LPRNTTSYVLRGLEPGQEYNVLLTAEKGRHKSKPARVKASTEQAPELENLTVTEVGWD
                     GLRLNWTAADQAYEHFIIQVQEANKVEAARNLTVPGSLRAVDIPGLKAATPYTVSIYG
                     VIQGYRTPVLSAEASTGETPNLGEVVVAEVGWDALKLNWTAPEGAYEYFFIQVQEADT
                     VEAAQNLTVPGGLRSTDLPGLKAATHYTITIRGVTQDFSTTPLSVEVLTEEVPDMGNL
                     TVTEVSWDALRLNWTTPDGTYDQFTIQVQEADQVEEAHNLTVPGSLRSMEIPGLRAGT
                     PYTVTLHGEVRGHSTRPLAVEVVTEDLPQLGDLAVSEVGWDGLRLNWTAADNAYEHFV
                     IQVQEVNKVEAAQNLTLPGSLRAVDIPGLEAATPYRVSIYGVIRGYRTPVLSAEASTA
                     KEPEIGNLNVSDITPESFNLSWMATDGIFETFTIEIIDSNRLLETVEYNISGAERTAH
                     ISGLPPSTDFIVYLSGLAPSIRTKTISATATTEALPLLENLTISDINPYGFTVSWMAS
                     ENAFDSFLVTVVDSGKLLDPQEFTLSGTQRKLELRGLITGIGYEVMVSGFTQGHQTKP
                     LRAEIVTEAEPEVDNLLVSDATPDGFRLSWTADEGVFDNFVLKIRDTKKQSEPLEITL
                     LAPERTRDLTGLREATEYEIELYGISKGRRSQTVSAIATTAMGSPKEVIFSDITENSA
                     TVSWRAPTAQVESFRITYVPITGGTPSMVTVDGTKTQTRLVKLIPGVEYLVSIIAMKG
                     FEESEPVSGSFTTALDGPSGLVTANITDSEALARWQPAIATVDSYVISYTGEKVPEIT
                     RTVSGNTVEYALTDLEPATEYTLRIFAEKGPQKSSTITAKFTTDLDSPRDLTATEVQS
                     ETALLTWRPPRASVTGYLLVYESVDGTVKEVIVGPDTTSYSLADLSPSTHYTAKIQAL
                     NGPLRSNMIQTIFTTIGLLYPFPKDCSQAMLNGDTTSGLYTIYLNGDKAQALEVFCDM
                     TSDGGGWIVFLRRKNGRENFYQNWKAYAAGFGDRREEFWLGLDNLNKITAQGQYELRV
                     DLRDHGETAFAVYDKFSVGDAKTRYKLKVEGYSGTAGDSMAYHNGRSFSTFDKDTDSA
                     ITNCALSYKGAFWYRNCHRVNLMGRYGDNNHSQGVNWFHWKGHEHSIQFAEMKLRPSN
                     FRNLEGRRKRA"
     polyA_signal    7522. .7527
BASE COUNT     1894 a    2019 c    2062 g    1585 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 9;  Length 7560;
  Best Local Similarity   83.3%;  Pred. No. 9.7e+02;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           | ||||||||||| || |||| |
Db    2462 CTTCAGGTGCAGGTAAGTACGTGG 2439


Search completed: December 26, 2001, 10:59:22
Job time: 2124 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:19:37 ; Search time 366.42 Seconds
(without alignments)
58.493 Million cell updates/sec

Title: US-09-396-196F-10
Perfect score: 25
Sequence: 1 tcgtcaggtgcaggtcagcacgttg 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	839	22	AAI62941	Human genomic DNA
	2	25	100.0	1041	20	AAX01303	E. coli biotin syn
	3	25	100.0	1084	10	AAN91329	E.coli Bio B gene.
	4	25	100.0	1121	7	AAN60496	Sequence encoding
	5	25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
c	6	17.6	70.4	7286	17	AAT14547	Cytotactin gene.
c	7	17.6	70.4	7346	20	AAX77924	Human tenascin cDN
c	8	17.6	70.4	7560	22	AAC83437	Human tenascin-C.
c	9	17.4	69.6	1215	22	AAH45073	Murine FM-3 coding
c	10	17.4	69.6	1526	21	AAZ49706	Mouse growth hormo
c	11	17.4	69.6	109519	22	AAS08693	Micromonospora DNA
c	12	17.2	68.8	92934	21	AAA81473	N. meningitidis pa
c	13	17.2	68.8	172325	21	AAF21613	Neisseria meningit
c	14	17.2	68.8	837096	21	AAA81489	N. meningitidis pa
	15	17	68.0	1155	22	AAH66024	C glutamicum codin
c	16	17	68.0	1572	22	AAH64967	C glutamicum codin
c	17	17	68.0	1883	14	AAQ45959	mRXR-alpha. Mus m
c	18	17	68.0	2095	12	AAQ13387	Mouse RXR-alpha co
c	19	17	68.0	3573	17	AAT36507	3.5 kb DNA contg.
c	20	17	68.0	20303	18	AAT71699	Human deoxycytidyl
c	21	17	68.0	26764	18	AAT71696	Human deoxycytidyl
c	22	17	68.0	349980	22	AAH64966	C glutamicum codin
	23	17	68.0	349980	22	AAH68527	C glutamicum codin
c	24	16.8	67.2	1183	20	AAZ31920	Human helicase, Re
c	25	16.8	67.2	1749	22	AAF87638	Human RecQ5 gamma
c	26	16.8	67.2	3080	22	AAD13357	Human secreted pro
c	27	16.8	67.2	3140	22	AAD13390	Human secreted pro
c	28	16.8	67.2	3703	22	AAF87636	Human RecQ5 beta D
c	29	16.8	67.2	3707	20	AAZ31910	Human helicase, Re
c	30	16.8	67.2	3715	22	AAF87637	Human RecQ5 alpha
	31	16.6	66.4	378	22	AAI13926	Probe #3859 for ge
	32	16.6	66.4	378	22	AAI35296	Probe #3982 used t
	33	16.6	66.4	378	22	AAI03795	Probe #3786 used t
	34	16.6	66.4	506	22	AAI42353	Probe #11039 used
	35	16.6	66.4	765	22	AAH68124	C glutamicum codin
	36	16.6	66.4	888	22	AAF71212	Corynebacterium gl
	37	16.6	66.4	888	22	AAF71213	Corynebacterium gl
c	38	16.6	66.4	891	20	AAZ16847	Human gene express
c	39	16.6	66.4	891	20	AAZ16835	Human gene express
c	40	16.6	66.4	1953	21	AAC77412	Human ORFX ORF2967
c	41	16.6	66.4	2414	22	AAH15118	Human cDNA sequenc
	42	16.6	66.4	309400	22	AAH68534	C glutamicum codin
	43	16.2	64.8	624	22	AAS05572	Mammalian vestibul
c	44	16.2	64.8	735	19	AAV52499	Streptococcus pneu
	45	16.2	64.8	1299	18	AAX30776	Streptococcus pneu
	46	16.2	64.8	1496	8	AAN71388	Encodes subtilisin
c	47	16.2	64.8	2517	20	AAZ28269	Rat neuronal immed
c	48	16.2	64.8	2674	19	AAZ96299	S. pneumoniae deri
c	49	16.2	64.8	2674	19	AAV42984	Streptococcus pneu
c	50	16.2	64.8	10357	19	AAV52324	Streptococcus pneu
	51	16.2	64.8	49272	19	AAV35000	Mycobacteriophage
c	52	16	64.0	258	21	AAZ46040	Partial cDNA encod
c	53	16	64.0	403	21	AAF12855	Aspergillus oryzae
c	54	16	64.0	450	22	AAI12043	Probe #1976 for ge
c	55	16	64.0	450	22	AAI33378	Probe #2064 used t
c	56	16	64.0	450	22	AAI01968	Probe #1959 used t
	57	16	64.0	562	22	AAH09614	Human cDNA clone (
c	58	16	64.0	1188	21	AAC43677	Zea mays DNA fragm
c	59	16	64.0	1866	12	AAQ13386	Human RXR-alpha co
	60	16	64.0	2502	18	AAT88991	Streptococcus pneu
c	61	16	64.0	2586	21	AAA79731	Eucalyptus grandis
	62	16	64.0	2895	22	AAH16308	Human cDNA sequenc
	63	16	64.0	3369	22	AAH14577	Human cDNA sequenc
	64	16	64.0	3978	22	AAH81312	Escherichia coli p
c	65	16	64.0	4376	21	AAA14818	DNA encoding a C1
c	66	16	64.0	5837	21	AAA92454	Murine E-selectin
	67	16	64.0	11770	19	AAV52305	Streptococcus pneu
	68	16	64.0	50925	21	AAA81487	N. meningitidis pa
	69	16	64.0	349980	21	AAF21610	Neisseria meningit
c	70	15.8	63.2	244	21	AAC30833	Human secreted pro
c	71	15.8	63.2	549	22	AAF71069	C. glutamicum SRT
	72	15.8	63.2	665	21	AAC43992	Arabidopsis thalia
	73	15.8	63.2	906	11	AAQ04701	Sequence encoding
	74	15.8	63.2	979	12	AAQ12149	Cytotoxic lymphocy
	75	15.8	63.2	984	18	AAT50759	Sequence of ovine
	76	15.8	63.2	987	16	AAQ97484	Human interleukin-
	77	15.8	63.2	987	19	AAV07273	Human Interleukin-
	78	15.8	63.2	987	19	AAV07217	DNA encoding human
	79	15.8	63.2	987	20	AAZ40404	Human interleukin
	80	15.8	63.2	987	20	AAV83199	Interleukin (IL)-1
	81	15.8	63.2	1018	15	AAQ67879	EPV 42 kD promoter
	82	15.8	63.2	1018	20	AAZ08510	EPV 42 kDa/human I
	83	15.8	63.2	1019	21	AAA70709	Bovine interleukin
c	84	15.8	63.2	1239	22	AAH45076	Rat FM-3 coding se

c	85	15.8	63.2	1272	22	AAH65106	C glutamicum codin
	86	15.8	63.2	1276	21	AAC45328	Arabidopsis thalia
c	87	15.8	63.2	1292	22	AAD08025	Rat G-protein coup
	88	15.8	63.2	1323	21	AAC45618	Arabidopsis thalia
	89	15.8	63.2	1325	21	AAC39434	Arabidopsis thalia
	90	15.8	63.2	1353	21	AAC45976	Arabidopsis thalia
	91	15.8	63.2	1353	21	AAC46402	Arabidopsis thalia
c	92	15.8	63.2	1360	22	AAF31248	Human immune respo
c	93	15.8	63.2	1368	22	AAF44623	Novel protein kina
c	94	15.8	63.2	1371	22	AAD08026	Rat G-protein coup
	95	15.8	63.2	1389	22	AAF76880	Human p40-IL-2 cod
c	96	15.8	63.2	1395	22	AAF71067	C. glutamicum SRT
c	97	15.8	63.2	1395	22	AAF68042	Corynebacterium gl
	98	15.8	63.2	1397	18	AAV00401	40 kDa subunit of
	99	15.8	63.2	1397	18	AAT77849	40 kDa subunit of
	100	15.8	63.2	1397	18	AAT97969	40 kDa subunit of

ALIGNMENTS

RESULT 1
AAI62941
ID AAI62941 standard; DNA; 839 BP.
XX
AC AAI62941;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human genomic DNA SEQ ID NO 269.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200155449-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01346.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUL-2000; 2000US-0216880.
PR 14-JUL-2000; 2000US-0218290.
PR 14-AUG-2000; 2000US-0225447.
PR 01-SEP-2000; 2000US-0229343.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231243.
PR 25-SEP-2000; 2000US-0234997.
PR 29-SEP-2000; 2000US-0236367.
PR 13-OCT-2000; 2000US-0239937.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246528.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249265.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-476225/51.
XX
PT Novel plasma membrane associated proteins useful for diagnosing,
PT treating, preventing and/or prognosing disorders related to the
PT proteins, including cancer, immune response and neuronal disorders -
XX
PS Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AAI62752-AAI62961) and proteins
CC (AAM42347-AAM42415) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;

Query Match 100.0%; Score 25; DB 22; Length 839;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcacgttg 25
||||||||||||||||||||||||||
Db 143 tcgtcaggtgcaggtcagcacgttg 167

RESULT 2
AAX01303
ID AAX01303 standard; DNA; 1041 BP.
XX
AC AAX01303;
XX
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (BioB) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS Escherichia coli.
XX
PN US5869719-A.
XX
PD 09-FEB-1999.
XX
PF 30-APR-1997; 97US-0846338.
XX
PR 30-APR-1997; 97US-0846338.
PR 08-MAR-1995; 95US-0401068.
XX

```
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Patton DA;
XX
DR   WPI; 1999-152902/13.
DR   P-PSDB; AAW73906.
XX
PT   Transgenic plants with high biotin levels - transformed with DNA
PT   encoding di:amino-pelargonic acid amino-transferase or biotin
PT   synthase
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match            100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.036;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    111 tcgtcaggtgcaggtcagcacgttg 135


RESULT   3
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA;
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match            100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity  100.0%;  Pred. No. 0.036;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    134 tcgtcaggtgcaggtcagcacgttg 158


RESULT   4
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzyme -
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match            100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity  100.0%;  Pred. No. 0.036;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    152 tcgtcaggtgcaggtcagcacgttg 176


RESULT   5
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
```

```
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry)
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag= a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag= b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag= c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag= d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag= e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag= f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag= g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag= h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag= i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag= j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag= k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag= l
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag= m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag= n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                   7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag= o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag= p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag= q
FT                   /standard_name= "rho-independent transcriptional
FT                                   terminator"
FT   stem_loop       5583..5605
FT                   /*tag= r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;   93WO-EP02688.
XX
PR   02-OCT-1992;   92CH-0003124.
PR   15-JUL-1993;   93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.043;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     227 tcgtcaggtgcaggtcagcacgttg 251


RESULT   6
AAT14547/c
ID   AAT14547 standard; DNA; 7286 BP.
XX
AC   AAT14547;
```

```
XX
DT   21-JUN-1996  (first entry)
XX
DE   Cytotactin gene.
XX
KW   Cytotactin; neuron; neurite; cell attachment; cell elongation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             55..6654
FT                   /*tag=  a
XX
PN   WO9608513-A1.
XX
PD   21-MAR-1996.
XX
PF   14-SEP-1995;   95WO-US11684.
XX
PR   16-SEP-1994;   94US-0308359.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   Crossin KL,  Phillips G,  Prieto AL;
XX
DR   WPI; 1996-179904/18.
DR   P-PSDB; AAR94562.
XX
PT   Cytotactin polypeptide(s), derivs. and antibodies - capable of
PT   stimulating neuronal cell attachment, neurite out-growth and cell
PT   elongation
XX
PS   Disclosure; Page 100-110; 159pp; English.
XX
CC   The human cytotactin gene (AAT14547) and chicken cytotactin gene
CC   (AAT14548) code for multidomain extracellular matrix proteins (AAR94562
CC   and AAR94563, respectively) capable of simulating neuronal cell
CC   attachment, neurite outgrowth and cell elongation.  The genes,
CC   or portions of them, can be inserted into a vector and used for
CC   prodn. in bacteria of e.g. fusion proteins between GST and CT.
XX
SQ   Sequence 7286 BP; 1830 A; 1926 C; 1999 G; 1531 T; 0 other;


  Query Match             70.4%;  Score 17.6;  DB 17;  Length 7286;
  Best Local Similarity   83.3%;  Pred. No. 90;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacgttg 25
          | ||||||||||||| || |||| |
Db   2200 CTTCAGGTGCAGGTAAGTACGTGG 2177


RESULT   7
AAX77924/c
ID   AAX77924 standard; cDNA; 7346 BP.
XX
AC   AAX77924;
XX
DT   16-AUG-1999  (first entry)
XX
DE   Human tenascin cDNA.
XX
KW   Tenascin; antipsoriasis; antivitiligo; anticancer; anti-inflammatory;
KW   cardiovascular; treatment; disease; depigmentation; albinism; cancer;
KW   psoriasis; vitiligo; metastasis; melanoma; inflammation; restenosis;
KW   diagnosis; human; ss.
XX
OS   Homo sapiens.
XX
PN   DE19750702-A1.
XX
PD   27-MAY-1999.
XX
PF   15-NOV-1997;   97DE-1050702.
XX
PR   15-NOV-1997;   97DE-1050702.
XX
PA   (HMRI ) HOECHST MARION ROUSSEL DEUT GMBH.
XX
PI   Peyman A,  Uhlmann E,  Weiser C;
XX
DR   WPI; 1999-314075/27.
XX
PT   Antisense oligonucleotides that bind to sequences encoding human
PT   tenascin  for treating depigmentation, cancer, inflammation and
PT   cardiovascular disease
XX
PS   Disclosure; Page 8-14; 18pp; German.
XX
CC   This invention describes novel oligonucleotides with up to 17 optionally
CC   modified nucleotides (nt), or their salts which are capable of binding
CC   to a nucleic acid encoding an isoform of human tenascin, or a part of it.
CC   The oligonucleotides of the invention have antipsoriasis, antivitiligo,
CC   anticancer, anti-inflammatory and cardiovascular activity. The
CC   oligonucleotides are used to treat or prevent diseases associated with
CC   (over)expression of tenascin, particularly depigmentation (albinism,
CC   psoriasis or vitiligo), cancer or metastases, particularly melanoma,
CC   inflammation or cardiovascular disease (e.g. restenosis). A preferred
CC   application is treatment of vitiligo. The oligonucleotides may also be
CC   used for diagnosis of these diseases.
XX
SQ   Sequence 7346 BP; 1840 A; 1945 C; 2020 G; 1541 T; 0 other;


  Query Match             70.4%;  Score 17.6;  DB 20;  Length 7346;
  Best Local Similarity   83.3%;  Pred. No. 90;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacgttg 25
          | ||||||||||||| || |||| |
Db   2200 CTTCAGGTGCAGGTAAGTACGTGG 2177


RESULT   8
AAC83437/c
ID   AAC83437 standard; DNA; 7560 BP.
XX
AC   AAC83437;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Human tenascin-C.
XX
KW   Tenascin-C; axon; dendrite; neurite; ds.
XX
OS   Homo sapiens.
XX
PN   WO200066628-A1.
XX
PD   09-NOV-2000.
XX
PF   28-APR-2000; 2000WO-US11647.
XX
PR   01-MAY-1999;   99US-0132137.
XX
PA   (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
PI   Geller HM,  Meiners S;
XX
DR   WPI; 2001-015974/02.
DR   P-PSDB; AAB36935.
XX
PT   New peptides and methods of stimulating axonal and/or dendritic growth
PT   and/or guidance by administering a peptide, particularly comprising a
PT   tenascin-C region, to a neuron  -
```

```
XX
PS   Disclosure; Page 11-16; 73pp; English.
XX
CC   The present invention relates to a tenascin-C peptide. The peptide,
CC   particularly from a tenascin-C region, is used for stimulating axonal
CC   and/or dendritic growth, independent of neurite guidance, and
CC   stimulating axonal and/or dendritic guidance independent of axonal
CC   and/or dendritic growth.
XX
SQ   Sequence 7560 BP; 1894 A; 2019 C; 2062 G; 1585 T; 0 other;


  Query Match             70.4%;  Score 17.6;  DB 22;  Length 7560;
  Best Local Similarity   83.3%;  Pred. No. 91;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacgttg 25
          | ||||||||||||  || |||| |
Db   2462 CTTCAGGTGCAGGTAAGTACGTGG 2439


RESULT   9
AAH45073/c
ID   AAH45073 standard; cDNA; 1215 BP.
XX
AC   AAH45073;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Murine FM-3 coding sequence.
XX
KW   Murine; FM-3; hypotensive; tranquiliser; neuromedin U; stress;
KW   hypertension; appetite regulator; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1215
FT                   /*tag=  a
FT                   /product= "Murine FM-3"
FT                   /note= "No stop codon given"
XX
PN   WO200140797-A1.
XX
PD   07-JUN-2001.
XX
PF   28-NOV-2000; 2000WO-JP08363.
XX
PR   29-NOV-1999;   99JP-0338387.
PR   24-FEB-2000; 2000JP-0052251.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Hinuma S,  Hosoya M,  Kawamata Y,  Fukusumi S;
XX
DR   WPI; 2001-374952/39.
DR   P-PSDB; AAB99186.
XX
PT   Method for screening for compounds that alter binding between
PT   neuromedin U and FM-3 useful for treatment of stress, hypertension and
PT   for regulating appetite  -
XX
PS   Disclosure; Page 76-78; 92pp; Japanese.
XX
CC   The present invention relates to a method for screening for compounds
CC   that alter binding between neuromedin U and its derivatives to FM-3 and
CC   its salts. The present sequence is the coding sequence for murine FM-3,
CC   which was used in the method of the present invention. The compounds
CC   found using the method of the present invention are useful for the
CC   treatment and prevention of stress, hypertension and also as an appetite
CC   regulator.
XX
SQ   Sequence 1215 BP; 238 A; 362 C; 333 G; 282 T; 0 other;


  Query Match             69.6%;  Score 17.4;  DB 22;  Length 1215;
  Best Local Similarity   94.7%;  Pred. No. 92;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagcacgttg 25
          |||||||||||||| |||||
Db    180 GGTGCAGGTCAGCCCGTTG 162


RESULT  10
AAZ49706/c
ID   AAZ49706 standard; DNA; 1526 BP.
XX
AC   AAZ49706;
XX
DT   18-APR-2000  (first entry)
XX
DE   Mouse growth hormone secretagogue related receptor DNA.
XX
KW   Mouse growth hormone secretagogue related receptor; GHSR-R;
KW   human GHSR-R gene; ligand; agonist; antagonist; GHSR-R agonist;
KW   neuromedin U; obesity; depression; anxiety; compulsion; neuroses;
KW   insomnia; sleep disorder; substance abuse; pain; neuroprotective;
KW   cognitive disorder; memory enhancement; Alzheimer's disease;
KW   GHSR-R antagonist; anorectic; nootropic; antidepressant; tranquilizer;
KW   antiaddictive; analgesic; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             309..1526
FT                   /*tag=  a
FT                   /product= "Mouse growth hormone secretagogue related
FT                   receptor"
XX
PN   WO200002919-A1.
XX
PD   20-JAN-2000.
XX
PF   13-JUL-1999;   99WO-US15941.
XX
PR   13-JUL-1998;   98US-0092623.
XX
PA   (MERI ) MERCK & CO INC.
XX
PI   Liu Q,  Howard AD,  Mckee KK;
XX
DR   WPI; 2000-147599/13.
DR   P-PSDB; AAY44641.
XX
PT   New receptors useful for identifying agonist and antagonist for
PT   treating or preventing obesity, anxiety, neuroses, insomnia  -
XX
PS   Claim 9; Fig 1; 50pp; English.
XX
CC   The present sequence encodes mouse growth hormone secretagogue related
CC   receptor (GHSR-R). This was isolated from a mouse EST derived from a
CC   T-cell library. This exhibits 63% identity with human GHSR-R gene. GHSR-R
CC   is useful for identifying ligands, agonist or antagonist which binds to
CC   it. A GHSR-R agonist or neuromedin U is used in the treatment of
CC   obesity, depression, anxiety, compulsion, neuroses, insomnia/sleep
CC   disorders, substance abuse, pain, neuroprotective and cognitive disorders
CC   and memory enhancement including Alzheimer's disease. GHSR-R antagonist
CC   is useful for increasing food intake of a mammal.
XX
SQ   Sequence 1526 BP; 301 A; 453 C; 395 G; 377 T; 0 other;


  Query Match             69.6%;  Score 17.4;  DB 21;  Length 1526;
  Best Local Similarity   94.7%;  Pred. No. 94;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy      7 ggtgcaggtcagcacgttg 25
          ||||||||||||| |||||
Db    488 GGTGCAGGTCAGCCCGTTG 470


RESULT  11
AAS08693/c
ID   AAS08693 standard; DNA; 109519 BP.
XX
AC   AAS08693;
XX
DT   26-SEP-2001  (first entry)
XX
DE   Micromonospora DNA encoding biosynthetic enzymes for Everninomycin.
XX
KW   Everninomicin; antibiotic; bottle-neck gene; orthomicin;
KW   fermentation; ds.
XX
OS   Micromonospora carbonacea var. africana.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (132..1382)
FT                   /*tag= a
FT                   /product= "EvdA"
FT   RBS             complement (1389..1394)
FT                   /*tag= b
FT   CDS             complement (1490..2611)
FT                   /*tag= c
FT                   /product= "EvdB"
FT   RBS             complement (2618..2622)
FT                   /*tag= d
FT   CDS             complement (2622..3860)
FT                   /*tag= e
FT                   /product= "EvdC"
FT   RBS             complement (3867..3870)
FT                   /*tag= f
FT   CDS             4143..5312
FT                   /*tag= g
FT                   /product= "EvdD"
FT   RBS             4134..4138
FT                   /*tag= h
FT   CDS             5309..6235
FT                   /*tag= i
FT                   /product= "EvdE"
FT   CDS             6232..7275
FT                   /*tag= j
FT                   /product= "EvdF"
FT   RBS             6226..6229
FT                   /*tag= k
FT   CDS             7272..8327
FT                   /*tag= l
FT                   /product= "EvdG"
FT   CDS             8342..9364
FT                   /*tag= m
FT                   /product= "EvdH"
FT   RBS             8333..8336
FT                   /*tag= n
FT   CDS             complement (9463..10224)
FT                   /*tag= o
FT                   /product= "EvdI"
FT   RBS             complement (10232..10235)
FT                   /*tag= p
FT   CDS             10424..11176
FT                   /*tag= q
FT                   /product= "EvdJ"
FT   CDS             12027..12455
FT                   /*tag= r
FT                   /product= "EvdK"
FT                   /partial
FT                   /note= "No start codon"
FT   CDS             complement (12108..13022)
FT                   /*tag= s
FT                   /product= "EvdL"
FT   RBS             complement (13027..13030)
FT                   /*tag= t
FT   CDS             complement (14410..15363)
FT                   /*tag= u
FT                   /product= "EvrA"
FT   RBS             complement (15369..15373)
FT                   /*tag= v
FT   CDS             complement (15380..16414)
FT                   /*tag= w
FT                   /product= "EvrB"
FT   CDS             complement 16419..17873
FT                   /*tag= x
FT                   /product= "EvrC"
FT   CDS             complement (17870..18934)
FT                   /*tag= y
FT                   /product= "EvrD"
FT   CDS             19374..20906
FT                   /*tag= z
FT                   /product= "EvrE"
FT   CDS             21064..22542
FT                   /*tag= aa
FT                   /product= "EvrF"
FT   RBS             21056..22542
FT                   /*tag= ab
FT   CDS             22748..24172
FT                   /*tag= ac
FT                   /product= "EvrG"
FT   RBS             22736..22740
FT                   /*tag= ad
FT   CDS             complement (24177..25223)
FT                   /*tag= ae
FT                   /product= "EvrH"
FT   RBS             complement (25230..25233)
FT                   /*tag= af
FT   CDS             25550..26626
FT                   /*tag= ag
FT                   /product= "EvrI"
FT   CDS             26685..30479
FT                   /*tag= ah
FT                   /product= "EvrJ"
FT   RBS             26672..26676
FT                   /*tag= ai
FT   CDS             complement (30557..31876)
FT                   /*tag= aj
FT                   /product= "EvrK"
FT   RBS             complement (31885..31888)
FT                   /*tag= ak
FT   CDS             complement (31941..32882)
FT                   /*tag= al
FT                   /product= "EvrL"
FT   CDS             complement (33167..34405)
FT                   /*tag= am
FT                   /product= "EvrM"
FT   RBS             complement (34414..34418
FT                   /*tag= an
FT   CDS             complement (34449..35210)
FT                   /*tag= ao
FT                   /product= "EvrN"
FT   RBS             complement (35219..35221)
FT                   /*tag= ap
FT   CDS             complement (35294..36238)
FT                   /*tag= aq
FT                   /product= "EvrO"
FT   CDS             complement (36235..36963)
FT                   /*tag= ar
FT                   /product= "EvrP"
FT   CDS             complement (36998..38026)
FT                   /*tag= as
FT                   /product= "EvrQ"
FT   CDS             complement (38072..38566)
FT                   /*tag= at
FT                   /product= "EvrR"
```

```
FT   CDS             complement (38892..40163)
FT                   /*tag= au
FT                   /product= "EvrS"
FT   CDS             complement (40216..40890)
FT                   /*tag= av
FT                   /product= "EvrT"
FT   RBS             complement (40899..40902)
FT                   /*tag= aw
FT   CDS             complement (40887..41576)
FT                   /*tag= ax
FT                   /product= "EvrU"
FT   CDS             complement (41679..42707)
FT                   /*tag= ay
FT                   /product= "EvrV"
FT   RBS             complement (42714..42717)
FT                   /*tag= az
FT   CDS             complement (42810..43799)
FT                   /*tag= ba
FT                   /product= "EvrW"
FT   RBS             complement (43807..43811)
FT                   /*tag= bb
FT   CDS             complement (43799..44866)
FT                   /*tag= bc
FT                   /product= "EvrX"
FT   CDS             complement (45014..45760)
FT                   /*tag= bd
FT                   /product= "EvrY"
FT   RBS             complement (45767..45770)
FT                   /*tag= be
FT   CDS             complement (45962..46714)
FT                   /*tag= bf
FT                   /product= "EvrZ"
FT   RBS             complement (45952..45956)
FT                   /*tag= bg
FT   CDS             complement (47156..49234)
FT                   /*tag= bh
FT                   /product= "EvsA"
FT   CDS             51627..52715
FT                   /*tag= bi
FT                   /product= "EvsB"
FT   RBS             51629..51622
FT                   /*tag= bj
FT   CDS             52889..53557
FT                   /*tag= bk
FT                   /product= "EvsC"
FT   CDS             53554..54207
FT                   /*tag= bl
FT                   /product= "EvbA"
FT   CDS             complement (54362..55117)
FT                   /*tag= bm
FT                   /product= "EvbB"
FT   RBS             complement (55125..55128)
FT                   /*tag= bn
FT   CDS             complement (55135..56094)
FT                   /*tag= bo
FT                   /product= "EvbC"
FT   RBS             complement (56100..56103)
FT                   /*tag= bp
FT   CDS             complement (56184..56813)
FT                   /*tag= bq
FT                   /product= "EvbC2"
FT   CDS             56961..58709

  Query Match             69.6%;  Score 17.4;  DB 22;  Length 109519;
  Best Local Similarity   94.7%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagcacgttg 25
          ||||||| ||||||||||
Db  77002 GGTGCAGATCAGCACGTTG 76984


RESULT  12
```

```
AAA81473/c
ID   AAA81473 standard; DNA; 92934 BP.
XX
AC   AAA81473;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_21 SEQ ID NO:21.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 471-498; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 92934 BP; 24051 A; 24249 C; 21902 G; 22730 T; 2 other;


  Query Match             68.8%;  Score 17.2;  DB 21;  Length 92934;
  Best Local Similarity   86.4%;  Pred. No. 1.8e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtcagcacgttg 25
          |||||||||  ||| |||||||
```

```
Db  60036 TCAGGTGCAATTCAACACGTTG 60015


RESULT  13
AAF21613/c
ID   AAF21613 standard; DNA; 172325 BP.
XX
AC   AAF21613;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Neisseria meningitidis B nucleotide sequence SEQ ID NO:114.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; immunogenic; vaccine;
KW   diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW   ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200066791-A1.
XX
PD   09-NOV-2000.
XX
PF   08-MAR-2000; 2000WO-US05928.
XX
PR   30-APR-1999;   99US-0132068.
PR   08-OCT-1999;   99WO-US23573.
PR   28-FEB-2000; 2000GB-0004695.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Pizza M,  Hickey E,  Peterson J,  Tettelin H,  Venter JC,  Masignani V;
PI   Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V,  Rappuoli R;
PI   Frazer CM,  Grandi G;
XX
DR   WPI; 2000-647603/62.
XX
PT   Neisseria meningitidis B full length genome sequence and open reading
PT   frames are used to detect, treat and prevent Neisserial infections -
XX
PS   Claim 7; Appendix A; 692pp; English.
XX
CC   The present invention describes the full length genome of
CC   Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC   to AAF21613 represent fragments of the NMB genomic sequence, as the
CC   sequence was too long to go in a record on its own it was split into 8
CC   sequences which overlap each other at the beginning and end of each
CC   sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC   the beginning  of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC   the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC   Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC   AAF21606 represent PCR primers which are used in the exemplification of
CC   the present invention. The NMB genome and fragments from it have
CC   antibacterial activity, and can be used in vaccines and gene therapy.
CC   Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC   proteins can be used in compositions for treating or preventing infection
CC   due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC   presence of Neisserial bacteria or of antibodies raised to Neisserial
CC   bacteria. Computers, computer memory, computer storage medium or computer
CC   databases can be used in a search to identify open reading frames (ORFs)
CC   or coding sequences within the NMB genome. The DNA sequences provide
CC   further opportunities to find antigenic or immunogenic proteins which are
CC   more effective in vaccines than the outer membrane proteins currently
CC   used.
XX
SQ   Sequence 172325 BP; 43072 A; 47583 C; 41465 G; 40205 T; 0 other;


  Query Match            68.8%;  Score 17.2;  DB 21;  Length 172325;
  Best Local Similarity  86.4%;  Pred. No. 1.9e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtcagcacgttg 25
           |||||||||  |||  |||||||
Db  163703 TCAGGTGCAATTCAACACGTTG 163682


RESULT  14
AAA81489/c
ID   AAA81489 standard; DNA; 837096 BP.
XX
AC   AAA81489;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_37 SEQ ID NO:37.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 629-865; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;
```

```
  Query Match          68.8%;  Score 17.2;  DB 21;  Length 837096;
  Best Local Similarity   86.4%;  Pred. No. 2e+02;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtcagcacgttg 25
           |||||||||  ||| |||||||
Db 828474 TCAGGTGCAATTCAACACGTTG 828453


RESULT  15
AAH66024
ID   AAH66024 standard; DNA; 1155 BP.
XX
AC   AAH66024;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 1059.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI   Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR   WPI; 2001-376931/40.
DR   P-PSDB; AAG90805.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT    -
XX
PS   Claim 8; SEQ ID NO: 1059; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence 1155 BP; 218 A; 304 C; 327 G; 306 T; 0 other;


  Query Match          68.0%;  Score 17;  DB 22;  Length 1155;
  Best Local Similarity   80.0%;  Pred. No. 1.4e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           ||| ||| ||||||||||  || ||
Db     418 tcggcagctgcaggtcaggtcggtg 442


RESULT  16
AAH64967/c
ID   AAH64967 standard; DNA; 1572 BP.
XX
AC   AAH64967;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 2.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI   Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR   WPI; 2001-376931/40.
DR   P-PSDB; AAG89748.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT    -
XX
PS   Claim 8; SEQ ID NO: 2; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence 1572 BP; 392 A; 451 C; 401 G; 328 T; 0 other;


  Query Match          68.0%;  Score 17;  DB 22;  Length 1572;
  Best Local Similarity   80.0%;  Pred. No. 1.4e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           ||||||||| ||  |  ||||||||
Db     133 TCGTCAGGTTCAAATATGCACGTTG 109


RESULT  17
AAQ45959/c
ID   AAQ45959 standard; DNA; 1883 BP.
XX
AC   AAQ45959;
```

```
XX
DT   01-FEB-1994  (first entry)
XX
DE   mRXR-alpha.
XX
KW   Retinoid X receptor; RXR; retinoic acid receptor; RAR;
KW   thyroid receptor; TR; retinoic acid response elements; RARE;
KW   thyroid receptor response element; TRE;
KW   retinoid X receptor response element; RXRE;
KW   subunit; heterodimer; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             96..1499
FT                   /*tag=  a
FT                   /label= RXR-beta
XX
PN   WO9315216-A.
XX
PD   05-AUG-1993.
XX
PF   25-JAN-1993;   93WO-US00639.
XX
PR   24-JAN-1992;   92US-0825667.
XX
PA   (CNRS ) CENT NAT RECH SCI.
PA   (INRM ) INST NAT SANTE & RECH MEDICALE.
PA   (SQUI ) SQUIBB & SONS INC E R.
PA   (UYST-) UNIV PASTEUR STRASBOURG LOUIS.
XX
PI   Chambon P,  Kastner P,  Leid M;
XX
DR   WPI; 1993-258691/32.
DR   P-PSDB; AAR39470.
XX
PT   Heterodimer comprising retinoic acid or thyroid receptor and
PT   retinoid X receptor - useful in diagnosis and for generation of
PT   antibodies
XX
PS   Claim 25; Page 97-99; 128pp; English.
XX
CC   Two different types of nuclear receptors, RAR and TR dimerise with
CC   RXR to form a heterodimer which is capable of binding to RARE, TRE
CC   or RXRE at physiological conditions.  Sequences encoding 1 isoform
CC   of mouse RXR-beta, 3 isoforms of human RXR-beta, 1 isoform of mouse
CC   RXR-alpha and 1 isoform of mouse RXR-gamma are provided (AAQ45955-60).
XX
SQ   Sequence 1883 BP; 380 A; 595 C; 518 G; 390 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 14;  Length 1883;
  Best Local Similarity   80.0%;  Pred. No. 1.5e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          ||| |||||| ||||||| | |||
Db    626 TCGGCAGGTGTAGGTCAGGTCTTTG 602


RESULT  18
AAQ13387/c
ID   AAQ13387 standard; cDNA; 2095 BP.
XX
AC   AAQ13387;
XX
DT   07-NOV-1991  (first entry)
XX
DE   Mouse RXR-alpha coding sequence.
XX
KW   ligand-binding domain receptor-alpha; retinoid receptor;
KW   vitamin A metabolite; hRAR-alpha; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             177..1581
FT                   /*tag=  a
FT                   /product= mRXR-alpha
XX
PN   WO9112258-A.
XX
PD   22-AUG-1991.
XX
PF   22-JAN-1991;   91WO-US00399.
XX
PR   09-FEB-1990;   90US-0478071.
XX
PA   (SALK ) SALK INST FOR BIOL STUD.
XX
PI   Mangelsdorf DJ,  Evans RM;
XX
DR   WPI; 1991-267086/36.
DR   P-PSDB; AAR13799.
XX
PT   Novel DNA sequence encoding a receptor polypeptide - which is
PT   useful for testing the ability of cpds. to regulate their
PT   transcription - activation effects
XX
PS   Claim 4; Page 29-30; 56pp; English.
XX
CC   mRXR-alpha is a novel retinoid receptor. Clones were isolated from a
CC   lambda gt11 human liver cDNA library by screening with the known
CC   hRAR-alpha DNA binding domain. These positive clones were subcloned
CC   and a unique receptor-like sequence was identified and used to
CC   screen a lambda gt10 human kidney cDNA library. A human RXR-alpha
CC   coding sequence was identified and this was used to identify
CC   mRXR-alpha from a mouse embryo library. Labelled single-stranded
CC   nucleic acid sequences are also claimed comprising 20 or more
CC   contiguous bases from within bases 20-2095 of the mRXR-alpha
CC   sequence. These can be used to identify other DNAs encoding
CC   retinoid receptors. See also AAQ13386 and AAQ13388.
XX
SQ   Sequence 2095 BP; 404 A; 680 C; 598 G; 413 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 12;  Length 2095;
  Best Local Similarity   80.0%;  Pred. No. 1.5e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          ||| |||||| ||||||| | |||
Db    708 TCGGCAGGTGTAGGTCAGGTCTTTG 684


RESULT  19
AAT36507/c
ID   AAT36507 standard; DNA; 3573 BP.
XX
AC   AAT36507;
XX
DT   04-OCT-1996  (first entry)
XX
DE   3.5 kb DNA contg. ars region, rpmH, dnaA and dnaN genes.
XX
KW   ars; dnaA; dnaN; rpmH; Brevibacterium flavum; low expression;
KW   autonomous replication sequence; plasmid copy number; ds.
XX
OS   Brevibacterium flavum.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (61..84)
FT                   /*tag=  a
FT                   /note= "start of rpmH gene"
FT   CDS             932..2506
FT                   /*tag=  b
```

```
FT                   /product= dnaA gene product
FT   CDS             3030
FT                   /*tag=  c
FT                   /note= "partial dnaN gene product"
XX
PN   JP08089259-A.
XX
PD   09-APR-1996.
XX
PF   05-DEC-1994;   94JP-0301140.
XX
PR   26-JUL-1994;   94JP-0174316.
XX
PA   (MITU ) MITSUBISHI CHEM CORP.
XX
DR   WPI; 1996-233351/24.
DR   P-PSDB; AAR98522-24.
XX
PT   DNA fragment contg. an autonomous replication sequence - useful for
PT   controlling the copy number of vectors used in cloning
XX
PS   Example 1; Page 11-14; 17pp; Japanese.
XX
CC   The present sequence is a 3.5 kb DNA fragment contg. an autonomous
CC   replicating sequence (ars) isolated from the coryneform bacterium,
CC   Brevibacterium flavum. The 3.5 kb BamHI-SalI fragment contains a partial
CC   rpmH gene, the dnaA gene and partial dnaN gene. The ars region was shown
CC   to be present between the dnaA and rpmH genes and between the dnaA and
CC   dnaN genes.  Vectors contg. the ars region have a low copy number which
CC   is useful for cloning of genes which adversely affect cells with high
CC   expression.
XX
SQ   Sequence 3573 BP; 932 A; 941 C; 872 G; 828 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 17;  Length 3573;
  Best Local Similarity   80.0%;  Pred. No. 1.6e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||| ||  |  |||||||
Db   1064 TCGTCAGGTTCAAATATGCACGTTG 1040


RESULT  20
AAT71699/c
ID   AAT71699 standard; DNA; 20303 BP.
XX
AC   AAT71699;
XX
DT   20-AUG-1997  (first entry)
XX
DE   Human deoxycytidylate deaminase intron 2 encoding DNA.
XX
KW   Recombinant deaminase; dCMP; ds.
XX
OS   Homo sapiens.
XX
PN   US5622851-A.
XX
PD   22-APR-1997.
XX
PF   10-JAN-1995;   95US-0370975.
XX
PR   10-JAN-1995;   95US-0370975.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Maley F,  Maley GR,  Weiner KXB;
XX
DR   WPI; 1997-244391/22.
XX
PT   DNA encoding human deoxycytidylate deaminase - for production of
PT   recombinant deaminase
XX
PS   Claim 2; Column 83-100; 58pp; English.
XX
CC   The present sequence encodes the human deoxycytidylate (dCMP)
CC   deaminase intron 2, which comprises 20303 base pairs from nucleotides
CC   1964-22266 of the dCMP deaminase sense strand. The dCMP deaminase gene
CC   contains a 5' untranslated region (including the promoter), 5 exons,
CC   4 introns and a 3' untranslated region (including the stop signals).
CC   The gene can be used to produce recombinant dCMP deaminase, which can
CC   be used to convert dCMP to dUMP. Also, the dCMP gene can be altered
CC   (removed or mutated) to alter DNA replication in cells, which may lead
CC   to mutagenesis.
XX
SQ   Sequence 20303 BP; 5454 A; 4115 C; 5052 G; 5682 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 18;  Length 20303;
  Best Local Similarity   80.0%;  Pred. No. 1.9e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          | ||  |||||||| |||||  ||||
Db    285 TGGTGTGGTGCAGGCCAGCAGGTTG 261


RESULT  21
AAT71696/c
ID   AAT71696 standard; DNA; 26764 BP.
XX
AC   AAT71696;
XX
DT   20-AUG-1997  (first entry)
XX
DE   Human deoxycytidylate deaminase gene.
XX
KW   Recombinant deaminase; dCMP; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..1317
FT                   /*tag=  a
FT                   /note= "5' untranslated region, including promotor"
FT   exon            1318..1425
FT                   /*tag=  b
FT                   /number= 1
FT   intron          1426..1827
FT                   /*tag=  c
FT                   /number= 1
FT   exon            1828..1963
FT                   /*tag=  d
FT                   /number= 2
FT   intron          1964..22266
FT                   /*tag=  e
FT                   /number= 2
FT   exon            22267..22383
FT                   /*tag=  f
FT                   /number= 3
FT   intron          22384..23740
FT                   /*tag=  g
FT                   /number= 3
FT   exon            23741..23837
FT                   /*tag=  h
FT                   /number= 4
FT   intron          23838..25391
FT                   /*tag=  i
FT                   /number= 4
FT   exon            25392..25467
FT                   /*tag=  j
FT                   /number= 5
FT   misc_feature    25468..26764
FT                   /*tag=  k
```

```
FT                   /note= "3' untranslated region"
XX
PN   US5622851-A.
XX
PD   22-APR-1997.
XX
PF   10-JAN-1995;   95US-0370975.
XX
PR   10-JAN-1995;   95US-0370975.
XX
PA   (HEAL-) HEALTH RES INC.
XX
PI   Maley F,  Maley GR,  Weiner KXB;
XX
DR   WPI; 1997-244391/22.
DR   P-PSDB; AAW18205.
XX
PT   DNA encoding human deoxycytidylate deaminase - for production of
PT   recombinant deaminase
XX
PS   Claim 3; Column 55-78; 58pp; English.
XX
CC   The present sequence encodes the human deoxycytidylate (dCMP)
CC   deaminase gene, which contains a 5' untranslated region (including
CC   the promoter), 5 exons, 4 introns and a 3' untranslated region
CC   (including the stop signals). The gene can be used to produce
CC   recombinant dCMP deaminase, which can be used to convert dCMP to dUMP.
CC   Also, the dCMP gene can be altered (removed or mutated) to alter DNA
CC   replication in cells, which may lead to mutagenesis.
XX
SQ   Sequence 26764 BP; 7079 A; 5521 C; 6539 G; 7625 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 18;  Length 26764;
  Best Local Similarity   80.0%;  Pred. No. 1.9e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          | ||  |||||||| ||||| ||||
Db   2248 TGGTGTGGTGCAGGCCAGCAGGTTG 2224


RESULT  22
AAH64966/c
ID   AAH64966 standard; DNA; 349980 BP.
XX
AC   AAH64966;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 1.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI   Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR   WPI; 2001-376931/40.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT   -
XX
PS   Claim 7; SEQ ID NO: 1; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence 349980 BP; 79703 A; 91547 C; 98381 G; 80349 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 22;  Length 349980;
  Best Local Similarity   80.0%;  Pred. No. 2.4e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          | ||  |||||||| ||||| ||||
Db    133 TCGTCAGGTTCAAATATGCACGTTG 109


RESULT  23
AAH68527
ID   AAH68527 standard; DNA; 349980 BP.
XX
AC   AAH68527;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 7062.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI   Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR   WPI; 2001-376931/40.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT   -
XX
PS   Disclosure; SEQ ID NO: 7062; 246pp + Sequence Listing; English.
XX
```

```
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence 349980 BP; 79725 A; 90426 C; 98918 G; 80911 T; 0 other;


  Query Match             68.0%;  Score 17;  DB 22;  Length 349980;
  Best Local Similarity   80.0%;  Pred. No. 2.4e+02;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          ||| ||| ||||||||||  || ||
Db 107415 tcggcagctgcaggtcaggtcggtg 107439


RESULT  24
AAZ31920/c
ID   AAZ31920 standard; cDNA to mRNA; 1183 BP.
XX
AC   AAZ31920;
XX
DT   25-JAN-2000  (first entry)
XX
DE   Human helicase, RecQ5, coding sequence fragment.
XX
KW   Helicase; RecQ5; human; homeostasis; cell aging; diagnosis; ss.
XX
OS   Homo sapiens.
XX
PN   JP11276173-A.
XX
PD   12-OCT-1999.
XX
PF   27-MAR-1998;   98JP-0081492.
XX
PR   27-MAR-1998;   98JP-0081492.
XX
PA   (EIJI-) EIJIN KENKYUSHO KK.
XX
DR   WPI; 1999-626937/54.
XX
PT   A gene coding helicase, RecQ5 - useful in research for elucidating the
PT   relationship to the maintenance of human homeostasis and cell aging
XX
PS   Example 1; Page 20; 25pp; Japanese.
XX
CC   This sequence is a fragment of the gene encoding the human helicase of
CC   the invention, designated RecQ5. The helicase RecQ5 gene can be used in a
CC   method for the diagnosis of diseases caused by an abnormality in the
CC   helicase gene. The gene can be used in researching the relationship
CC   between the maintenance of human homeostasis and cell aging.
XX
SQ   Sequence 1183 BP; 258 A; 313 C; 324 G; 288 T; 0 other;


  Query Match             67.2%;  Score 16.8;  DB 20;  Length 1183;
  Best Local Similarity   90.0%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagca 20
          || |||||||||| |||||
Db    713 TCTTCAGGTGCAGGGCAGCA 694


RESULT  25
AAF87638/c
ID   AAF87638 standard; cDNA; 1749 BP.
XX
AC   AAF87638;
XX
DT   16-JUL-2001  (first entry)
XX
DE   Human RecQ5 gamma DNA helicase cDNA.
XX
KW   Human; RecQ5 alpha; RecQ5 beta; RecQ5 gamma; DNA helicase;
KW   alternative splicing; chromosomal instability; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             161..1468
FT                   /*tag= a
FT                   /product= "RecQ5 gamma DNA helicase"
XX
PN   WO200125425-A1.
XX
PD   12-APR-2001.
XX
PF   25-AUG-2000; 2000WO-JP05757.
XX
PR   05-OCT-1999;   99JP-0284001.
XX
PA   (AGEN-) AGENE RES INST CO LTD.
XX
PI   Furuichi Y,  Shimamoto A,  Kitao S,  Nishikawa K;
XX
DR   WPI; 2001-273577/28.
DR   P-PSDB; AAB83197.
XX
PT   Polynucleotide encoding for RecQ5beta helicase useful for diagnosis and
PT   treatment of chromosomal instability -
XX
PS   Example 2; Page 72-75; 97pp; Japanese.
XX
CC   The present sequence is a polynucleotide encoding a human RecQ5
CC   type DNA helicase. The three RecQ5 type helicases alpha, beta and gamma
CC   are formed by alternative splicing. The invention discloses the RecQ5
CC   type DNA helicases beta and gamma, and the genes encoding them. The
CC   RecQ5 beta DNA helicase has a novel characteristic of being
CC   localised in the nucleus. It is useful as a diagnostic marker or in
CC   the treatment of diseases associated with chromosomal instability.
XX
SQ   Sequence 1749 BP; 406 A; 430 C; 486 G; 427 T; 0 other;


  Query Match             67.2%;  Score 16.8;  DB 22;  Length 1749;
  Best Local Similarity   90.0%;  Pred. No. 1.8e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagca 20
          || |||||||||| |||||
Db    791 TCTTCAGGTGCAGGGCAGCA 772


RESULT  26
AAD13357/c
ID   AAD13357 standard; cDNA; 3080 BP.
XX
AC   AAD13357;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 13 cDNA clone HSUMA53, SEQ ID NO:23.
XX
KW   Human secreted protein; proliferative disorder; cancer; tumour;
```

```
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             94..915
FT                   /*tag=  a
FT                   /product= "Human secreted protein precursor"
FT   sig_peptide     94..204
FT                   /*tag=  b
FT   mat_peptide     205..912
FT                   /*tag=  c
FT                   /product= "Mature human secreted protein"
XX
PN   WO200154708-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01434.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   18-AUG-2000; 2000US-0226279.
PR   05-DEC-2000; 2000US-0251988.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Komatsoulis GA,  Baker KP,  Birse CE,  Soppet DR,  Olsen HS;
PI   Moore PA,  Wei P,  Ebner R,  Duan DR,  Shi Y,  Choi GH,  Fiscella M;
PI   Ni J,  Ruben SM,  Barash SC;
XX
DR   WPI; 2001-488743/53.
DR   P-PSDB; AAE07063.
XX
PT   New isolated nucleic acids and polypeptides, useful for diagnosing,
PT   treating and/or preventing human diseases and disorders -
XX
PS   Claim 1; Page 455-456; 558pp; English.
XX
CC   AAD13345-AAD13401 represent cDNAs corresponding to 22 human secreted
CC   protein genes, and AAE07051-AAE07105 represent the proteins they encode.
CC   AAE07106-AAE07129 represent human secreted protein fragments or variants.
CC   The genes and their secreted proteins are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   22 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein-encoding cDNA of the invention.
XX
SQ   Sequence 3080 BP; 817 A; 629 C; 635 G; 983 T; 16 other;


  Query Match             67.2%;  Score 16.8;  DB 22;  Length 3080;
  Best Local Similarity   90.0%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagcacgttg 25
           |||| |||||||||| ||||
Db     430 AGGTCCAGGTCAGCAGGTTG 411


RESULT  27
AAD13390/c
ID   AAD13390 standard; cDNA; 3140 BP.
XX
AC   AAD13390;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 13 cDNA clone HSUMA53, SEQ ID NO:56.
XX
KW   Human secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             94..915
FT                   /*tag=  a
FT                   /product= "Human secreted protein precursor"
FT   sig_peptide     94..204
FT                   /*tag=  b
FT   mat_peptide     205..912
FT                   /*tag=  c
FT                   /product= "Mature human secreted protein"
XX
PN   WO200154708-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01434.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   18-AUG-2000; 2000US-0226279.
PR   05-DEC-2000; 2000US-0251988.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Komatsoulis GA,  Baker KP,  Birse CE,  Soppet DR,  Olsen HS;
PI   Moore PA,  Wei P,  Ebner R,  Duan DR,  Shi Y,  Choi GH,  Fiscella M;
PI   Ni J,  Ruben SM,  Barash SC;
XX
DR   WPI; 2001-488743/53.
DR   P-PSDB; AAE07095.
XX
```

```
PT   New isolated nucleic acids and polypeptides, useful for diagnosing,
PT   treating and/or preventing human diseases and disorders -
XX
PS   Claim 1; Page 480; 558pp; English.
XX
CC   AAD13345-AAD13401 represent cDNAs corresponding to 22 human secreted
CC   protein genes, and AAE07051-AAE07105 represent the proteins they encode.
CC   AAE07106-AAE07129 represent human secreted protein fragments or variants.
CC   The genes and their secreted proteins are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   22 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein-encoding cDNA of the invention.
XX
SQ   Sequence 3140 BP; 828 A; 645 C; 647 G; 1003 T; 17 other;


  Query Match             67.2%;  Score 16.8;  DB 22;  Length 3140;
  Best Local Similarity   90.0%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagcacgttg 25
           |||| |||||||||| ||||
Db     430 AGGTCCAGGTCAGCAGGTTG 411


RESULT  28
AAF87636/c
ID   AAF87636 standard; cDNA; 3703 BP.
XX
AC   AAF87636;
XX
DT   16-JUL-2001  (first entry)
XX
DE   Human RecQ5 beta DNA helicase cDNA.
XX
KW   Human; RecQ5 alpha; RecQ5 beta; RecQ5 gamma; DNA helicase;
KW   alternative splicing; chromosomal instability; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             161..3136
FT                   /*tag=  a
FT                   /product= "RecQ5 beta DNA helicase"
XX
PN   WO200125425-A1.
XX
PD   12-APR-2001.
XX
PF   25-AUG-2000; 2000WO-JP05757.
XX
PR   05-OCT-1999;   99JP-0284001.
XX
PA   (AGEN-) AGENE RES INST CO LTD.
XX
PI   Furuichi Y,  Shimamoto A,  Kitao S,  Nishikawa K;
XX
DR   WPI; 2001-273577/28.
DR   P-PSDB; AAB83195.
XX
PT   Polynucleotide encoding for RecQ5beta helicase useful for diagnosis and
PT   treatment of chromosomal instability -
XX
PS   Claim 1; Page 56-62; 97pp; Japanese.
XX
CC   The present sequence is a polynucleotide encoding a human RecQ5
CC   type DNA helicase. The three RecQ5 type helicases alpha, beta and gamma
CC   are formed by alternative splicing. The invention discloses the RecQ5
CC   type DNA helicases beta and gamma, and the genes encoding them. The
CC   RecQ5 beta DNA helicase has a novel characteristic of being
CC   localised in the nucleus. It is useful as a diagnostic marker or in
CC   the treatment of diseases associated with chromosomal instability.
XX
SQ   Sequence 3703 BP; 818 A; 1073 C; 1096 G; 716 T; 0 other;


  Query Match             67.2%;  Score 16.8;  DB 22;  Length 3703;
  Best Local Similarity   90.0%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagca 20
           || |||||||||| |||||
Db     791 TCTTCAGGTGCAGGGCAGCA 772


RESULT  29
AAZ31910/c
ID   AAZ31910 standard; cDNA; 3707 BP.
XX
AC   AAZ31910;
XX
DT   25-JAN-2000  (first entry)
XX
DE   Human helicase, RecQ5, coding sequence.
XX
KW   Helicase; RecQ5; human; homeostasis; cell aging; diagnosis; ss.
XX
OS   Homo sapiens.
XX
PN   JP11276173-A.
XX
PD   12-OCT-1999.
XX
PF   27-MAR-1998;   98JP-0081492.
XX
PR   27-MAR-1998;   98JP-0081492.
XX
PA   (EIJI-) EIJIN KENKYUSHO KK.
XX
DR   WPI; 1999-626937/54.
DR   P-PSDB; AAY43327.
XX
PT   A gene coding helicase, RecQ5 - useful in research for elucidating the
PT   relationship to the maintenance of human homeostasis and cell aging
XX
PS   Claim 2; Page 13-15; 25pp; Japanese.
XX
CC   This sequence encodes the human helicase of the invention, designated
CC   RecQ5. The helicase RecQ5 gene can be used in a method for the diagnosis
CC   of diseases caused by an abnormality in the helicase gene. The gene can
CC   be used in researching the relationship between the maintenance of human
CC   homeostasis and cell aging.
XX
SQ   Sequence 3707 BP; 801 A; 915 C; 1021 G; 970 T; 0 other;
```

```
  Query Match            67.2%;  Score 16.8;  DB 20;  Length 3707;
  Best Local Similarity  90.0%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagca 20
          || |||||||||| |||||
Db    791 TCTTCAGGTGCAGGGCAGCA 772


RESULT  30
AAF87637/c
ID   AAF87637 standard; cDNA; 3715 BP.
XX
AC   AAF87637;
XX
DT   16-JUL-2001  (first entry)
XX
DE   Human RecQ5 alpha DNA helicase cDNA.
XX
KW   Human; RecQ5 alpha; RecQ5 beta; RecQ5 gamma; DNA helicase;
KW   alternative splicing; chromosomal instability; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             161..3193
FT                   /*tag=  a
FT                   /product= "RecQ5 alpha DNA helicase"
XX
PN   WO200125425-A1.
XX
PD   12-APR-2001.
XX
PF   25-AUG-2000; 2000WO-JP05757.
XX
PR   05-OCT-1999;   99JP-0284001.
XX
PA   (AGEN-) AGENE RES INST CO LTD.
XX
PI   Furuichi Y,  Shimamoto A,  Kitao S,  Nishikawa K;
XX
DR   WPI; 2001-273577/28.
DR   P-PSDB; AAB83196.
XX
PT   Polynucleotide encoding for RecQ5beta helicase useful for diagnosis and
PT   treatment of chromosomal instability -
XX
PS   Example 2; Page 66-70; 97pp; Japanese.
XX
CC   The present sequence is a polynucleotide encoding a human RecQ5
CC   type DNA helicase. The three RecQ5 type helicases alpha, beta and gamma
CC   are formed by alternative splicing. The invention discloses the RecQ5
CC   type DNA helicases beta and gamma, and the genes encoding them. The
CC   RecQ5 beta DNA helicase has a novel characteristic of being
CC   localised in the nucleus. It is useful as a diagnostic marker or in
CC   the treatment of diseases associated with chromosomal instability.
XX
SQ   Sequence 3715 BP; 807 A; 916 C; 1021 G; 971 T; 0 other;


  Query Match            67.2%;  Score 16.8;  DB 22;  Length 3715;
  Best Local Similarity  90.0%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagca 20
          || |||||||||| |||||
Db    791 TCTTCAGGTGCAGGGCAGCA 772


RESULT  31
AAI13926
ID   AAI13926 standard; DNA; 378 BP.
XX
AC   AAI13926;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #3859 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 3859; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 378 BP; 92 A; 124 C; 89 G; 73 T; 0 other;


  Query Match            66.4%;  Score 16.6;  DB 22;  Length 378;
  Best Local Similarity  82.6%;  Pred. No. 1.9e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcagcacgttg 25
          | |||||| ||||||||| || |
Db     68 gacaggtgtaggtcagcaggtgg 90


RESULT  32
AAI35296
ID   AAI35296 standard; DNA; 378 BP.
XX
AC   AAI35296;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #3982 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
```

```
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 3982; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 378 BP; 92 A; 124 C; 89 G; 73 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 22;  Length 378;
  Best Local Similarity   82.6%;  Pred. No. 1.9e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           | |||||| |||||||| || |
Db      68 gacaggtgtaggtcagcaggtgg 90


RESULT  33
AAI03795
ID   AAI03795 standard; DNA; 378 BP.
XX
AC   AAI03795;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Probe #3786 used to measure gene expression in human breast sample.
XX
KW   Probe; human; breast disease; breast cancer; development disorder; ss;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 25; SEQ ID No 3786; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes.
CC   The present sequence is one such probe. The probes are useful for
CC   measuring human gene expression in a human breast sample, where the probe
CC   hybridises at high stringency to a nucleic acid expressed in the human
CC   breast. The probes are useful for predicting, diagnosing, grading,
CC   staging, monitoring and prognosing diseases of the human breast,
CC   particularly those diseases with polygenic aetiology. The diseases
CC   include: breast cancer, disorders of development, inflammatory diseases
CC   of the breast, fibrocystic changes, proliferative breast disease and
CC   non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 378 BP; 92 A; 124 C; 89 G; 73 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 22;  Length 378;
  Best Local Similarity   82.6%;  Pred. No. 1.9e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           | |||||| |||||||| || |
Db      68 gacaggtgtaggtcagcaggtgg 90


RESULT  34
AAI42353
ID   AAI42353 standard; DNA; 506 BP.
XX
AC   AAI42353;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #11039 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
```

```
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 11039; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 506 BP; 120 A; 147 C; 146 G; 93 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 22;  Length 506;
  Best Local Similarity   82.6%;  Pred. No. 1.9e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           | |||||| ||||||||| || |
Db      13 gacaggtgtaggtcagcaggtgg 35


RESULT  35
AAH68124
ID   AAH68124 standard; DNA; 765 BP.
XX
AC   AAH68124;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 3159.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI   Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR   WPI; 2001-376931/40.
DR   P-PSDB; AAG92905.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT    -
XX
PS   Claim 8; SEQ ID NO: 3159; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence 765 BP; 184 A; 199 C; 216 G; 166 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 22;  Length 765;
  Best Local Similarity   82.6%;  Pred. No. 2e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           | |||||| || || ||||||||
Db     552 gccaggtggagttctgcacgttg 574


RESULT  36
AAF71212
ID   AAF71212 standard; DNA; 888 BP.
XX
AC   AAF71212;
XX
DT   30-APR-2001  (first entry)
XX
DE   Corynebacterium glutamicum HA protein nucleotide sequence SEQ ID NO:149.
XX
KW   Corynebacterium glutamicum; homeostasis; adaptation; HA protein;
KW   fine chemical production; organic acid; proteinogenic amino acid;
KW   nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW   nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW   carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW   diagnosis; Corynebacterium diphtheriae; genetic engineering;
KW   Brevibacterium; environmental condition; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   WO200100842-A2.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-IB00911.
XX
PR   25-JUN-1999;   99US-0141031.
PR   08-JUL-1999;   99DE-1031636.
PR   09-JUL-1999;   99DE-1032125.
PR   09-JUL-1999;   99DE-1032126.
PR   09-JUL-1999;   99DE-1032127.
PR   09-JUL-1999;   99DE-1032128.
PR   09-JUL-1999;   99DE-1032129.
PR   09-JUL-1999;   99DE-1032226.
PR   14-JUL-1999;   99DE-1032920.
PR   14-JUL-1999;   99DE-1032922.
PR   14-JUL-1999;   99DE-1032924.
PR   14-JUL-1999;   99DE-1032928.
PR   14-JUL-1999;   99DE-1032930.
PR   14-JUL-1999;   99DE-1032933.
PR   14-JUL-1999;   99DE-1032935.
PR   14-JUL-1999;   99DE-1032973.
PR   14-JUL-1999;   99DE-1033002.
PR   14-JUL-1999;   99DE-1033003.
PR   14-JUL-1999;   99DE-1033005.
PR   14-JUL-1999;   99DE-1033006.
PR   31-AUG-1999;   99DE-1041378.
PR   31-AUG-1999;   99DE-1041379.
PR   31-AUG-1999;   99DE-1041390.
PR   31-AUG-1999;   99DE-1041391.
PR   03-SEP-1999;   99DE-1042088.
XX
PA   (BADI ) BASF AG.
XX
PI   Pompejus M,  Kroeger B,  Schroeder H,  Zelder O,  Haberhauer G;
XX
```

```
DR   WPI; 2001-061974/07.
DR   P-PSDB; AAB79097.
XX
PT   New isolated Corynebacterium glutamicum nucleic acid for production or
PT   modulation of production of fine chemicals such as amino acids,
PT   nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins
PT   or enzymes -
XX
PS   Claim 3; Page 353-354; 712pp; English.
XX
CC   AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis
CC   and adaptation (HA) proteins given in AAB79023 to AAB79242. The
CC   C. glutamicum HA genes (I) can be used in vectors for expression in host
CC   cells and production of fine chemicals, such as, an organic acid,
CC   proteinogenic or nonproteinogenic amino acid (preferred), purine or
CC   pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated
CC   fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,
CC   polyketide or enzyme. The amino acids produced can be lysine, glutamine,
CC   glutamate, alanine, aspartate, glycine, serine, threonine, methionine,
CC   cysteine, valine, leucine, isoleucine, arginine, proline, histidine,
CC   tyrosine, phenylalanine, or tryptophan. The fine chemical production can
CC   be modulated. The presence of (I) or HA proteins encoded by then are
CC   used for diagnosing the presence or activity of Corynebacterium
CC   diphtheriae. (I) can be used to map the C. glutamicum genome or can be
CC   used as markers for genetically engineered Corynebacterium or
CC   Brevibacterium. The HA proteins encoded by the (I) are used to maintain
CC   homeostasis in C. glutamicum or help the microorganism to adapt to
CC   different environmental conditions.
XX
SQ   Sequence 888 BP; 229 A; 217 C; 244 G; 198 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 22;  Length 888;
  Best Local Similarity   82.6%;  Pred. No. 2e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           | |||||| || || ||||||||
Db     652 gccaggtggagttctgcacgttg 674


RESULT  37
AAF71213
ID   AAF71213 standard; DNA; 888 BP.
XX
AC   AAF71213;
XX
DT   30-APR-2001  (first entry)
XX
DE   Corynebacterium glutamicum HA protein nucleotide sequence SEQ ID NO:151.
XX
KW   Corynebacterium glutamicum; homeostasis; adaptation; HA protein;
KW   fine chemical production; organic acid; proteinogenic amino acid;
KW   nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW   nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW   carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW   diagnosis; Corynebacterium diphtheriae; genetic engineering;
KW   Brevibacterium; environmental condition; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   WO200100842-A2.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-IB00911.
XX
PR   25-JUN-1999;   99US-0141031.
PR   08-JUL-1999;   99DE-1031636.
PR   09-JUL-1999;   99DE-1032125.
PR   09-JUL-1999;   99DE-1032126.
PR   09-JUL-1999;   99DE-1032127.
PR   09-JUL-1999;   99DE-1032128.
PR   09-JUL-1999;   99DE-1032129.
PR   09-JUL-1999;   99DE-1032226.
PR   14-JUL-1999;   99DE-1032920.
PR   14-JUL-1999;   99DE-1032922.
PR   14-JUL-1999;   99DE-1032924.
PR   14-JUL-1999;   99DE-1032928.
PR   14-JUL-1999;   99DE-1032930.
PR   14-JUL-1999;   99DE-1032933.
PR   14-JUL-1999;   99DE-1032935.
PR   14-JUL-1999;   99DE-1032973.
PR   14-JUL-1999;   99DE-1033002.
PR   14-JUL-1999;   99DE-1033003.
PR   14-JUL-1999;   99DE-1033005.
PR   14-JUL-1999;   99DE-1033006.
PR   31-AUG-1999;   99DE-1041378.
PR   31-AUG-1999;   99DE-1041379.
PR   31-AUG-1999;   99DE-1041390.
PR   31-AUG-1999;   99DE-1041391.
PR   03-SEP-1999;   99DE-1042088.
XX
PA   (BADI ) BASF AG.
XX
PI   Pompejus M,  Kroeger B,  Schroeder H,  Zelder O,  Haberhauer G;
XX
DR   WPI; 2001-061974/07.
DR   P-PSDB; AAB79098.
XX
PT   New isolated Corynebacterium glutamicum nucleic acid for production or
PT   modulation of production of fine chemicals such as amino acids,
PT   nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins
PT   or enzymes -
XX
PS   Claim 3; Page 355-356; 712pp; English.
XX
CC   AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis
CC   and adaptation (HA) proteins given in AAB79023 to AAB79242. The
CC   C. glutamicum HA genes (I) can be used in vectors for expression in host
CC   cells and production of fine chemicals, such as, an organic acid,
CC   proteinogenic or nonproteinogenic amino acid (preferred), purine or
CC   pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated
CC   fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,
CC   polyketide or enzyme. The amino acids produced can be lysine, glutamine,
CC   glutamate, alanine, aspartate, glycine, serine, threonine, methionine,
CC   cysteine, valine, leucine, isoleucine, arginine, proline, histidine,
CC   tyrosine, phenylalanine, or tryptophan. The fine chemical production can
CC   be modulated. The presence of (I) or HA proteins encoded by then are
CC   used for diagnosing the presence or activity of Corynebacterium
CC   diphtheriae. (I) can be used to map the C. glutamicum genome or can be
CC   used as markers for genetically engineered Corynebacterium or
CC   Brevibacterium. The HA proteins encoded by the (I) are used to maintain
CC   homeostasis in C. glutamicum or help the microorganism to adapt to
CC   different environmental conditions.
XX
SQ   Sequence 888 BP; 229 A; 217 C; 244 G; 198 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 22;  Length 888;
  Best Local Similarity   82.6%;  Pred. No. 2e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           | |||||| || || ||||||||
Db     652 gccaggtggagttctgcacgttg 674


RESULT  38
AAZ16847/c
ID   AAZ16847 standard; cDNA; 891 BP.
XX
AC   AAZ16847;
XX
DT   12-OCT-1999  (first entry)
XX
```

DE Human gene expression product cDNA sequence SEQ ID NO:4317.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
PS Claim 1; Page 2044; 2479pp; English.
XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 891 BP; 205 A; 188 C; 170 G; 191 T; 137 other;

Query Match 66.4%; Score 16.6; DB 20; Length 891;
Best Local Similarity 79.2%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcacgtt 24
||||||| ||| | ||||| |||
Db 169 TCGTCAGCAGCANGCCAGCAGGTT 146

RESULT 39
AAZ16835/c
ID AAZ16835 standard; cDNA; 891 BP.
XX
AC AAZ16835;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:4305.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
PS Claim 1; Page 2039; 2479pp; English.
XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 891 BP; 205 A; 188 C; 170 G; 191 T; 137 other;

Query Match 66.4%; Score 16.6; DB 20; Length 891;
Best Local Similarity 79.2%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcacgtt 24
||||||| ||| | ||||| |||
Db 169 TCGTCAGCAGCANGCCAGCAGGTT 146

```
RESULT  40
AAC77412/c
ID   AAC77412 standard; cDNA; 1953 BP.
XX
AC   AAC77412;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF2967 polynucleotide sequence SEQ ID NO:5933.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB43203.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 5116-5117; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 1953 BP; 442 A; 635 C; 554 G; 322 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 21;  Length 1953;
  Best Local Similarity   82.6%;  Pred. No. 2.2e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgtt 24
           || ||||||| ||||||| || |
Db     731 CGGCAGGTGCGGGTCAGCCCGCT 709


RESULT  41
AAH15118/c
ID   AAH15118 standard; cDNA; 2414 BP.
XX
AC   AAH15118;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:13158.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 13158; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
```

```
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2414 BP; 522 A; 732 C; 673 G; 487 T; 0 other;


  Query Match             66.4%;  Score 16.6;  DB 22;  Length 2414;
  Best Local Similarity   82.6%;  Pred. No. 2.3e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           | ||||||  ||||||||| || |
Db    1276 GACAGGTGTAGGTCAGCAGGTGG 1254


RESULT  42
AAH68534
ID   AAH68534 standard; DNA; 309400 BP.
XX
AC   AAH68534;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 7069.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI   Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR   WPI; 2001-376931/40.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT   -
XX
PS   Disclosure; SEQ ID NO: 7069; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence 309400 BP; 70133 A; 86477 C; 83115 G; 69675 T; 0 other;

  Query Match             66.4%;  Score 16.6;  DB 22;  Length 309400;
  Best Local Similarity   82.6%;  Pred. No. 3.5e+02;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           | ||||||  ||||||||| ||
Db   56418 gccaggtggagtt tgcacgttg 56440


RESULT  43
AAS05572
ID   AAS05572 standard; DNA; 624 BP.
XX
AC   AAS05572;
XX
DT   07-SEP-2001  (first entry)
XX
DE   Mammalian vestibular system geotactic behaviour modulator gene #172.
XX
KW   Mammalian vestibular system; invertebrate; geotactic behaviour; vertigo;
KW   graviperceptive disorder; motion sickness; labyrinthitis; syphilis; ds;
KW   Meniere's disease; acoustic neuroma; multiple sclerosis; epilepsy;
KW   trauma; infection of the middle ear; ototoxic agent exposure.
XX
OS   Drosophila melanogaster .
XX
PN   WO200140519-A2.
XX
PD   07-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-US32639.
XX
PR   02-DEC-1999;   99US-0168579.
PR   26-SEP-2000; 2000US-0669751.
XX
PA   (NEUR-) NEUROSCIENCES RES FOUND INC.
XX
PI   Greenspan RJ;
XX
DR   WPI; 2001-356159/37.
XX
PT   New isolated nucleic acid having mammalian vestibular system-modulating
PT   activity useful in the treatment of disorders such as motion sickness
PT   and vertigo -
XX
PS   Claim 59; Page 148-149; 179pp; English.
XX
CC   The sequences shown in AAS05401-AAS05661 represent DNA with mammalian
CC   vestibular system-modulating activity. The DNA sequences can be used in a
CC   method whereby a first and second strain of an invertebrate is obtained,
CC   and both are subjected to conditions in which the strains exhibit
CC   different geotactic behaviour. Genes that are differentially expressed in
CC   the first strain relative to the second strain are then identified.
CC   Mammalian genes having substantially the same nucleic acid sequence as
CC   these modulate the mammalian vestibular system. Compounds containing
CC   these genes are used to decrease the symptoms of graviperceptive
CC   disorders such as motion sickness, vertigo, labyrinthitis, Meniere's
CC   disease, acoustic neuroma, multiple sclerosis, syphilis, trauma,
CC   infection of the middle ear, exposure to ototoxic agents and epilepsy.
XX
SQ   Sequence 624 BP; 176 A; 145 C; 187 G; 116 T; 0 other;


  Query Match             64.8%;  Score 16.2;  DB 22;  Length 624;
  Best Local Similarity   85.7%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcac 21
           |||||||  |||||  ||||  ||
Db     537 tcgtcagttgcagatcagtac 557


RESULT  44
AAV52499/c
```

```
ID   AAV52499 standard; DNA; 735 BP.
XX
AC   AAV52499;
XX
DT   23-OCT-1998  (first entry)
XX
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:366.
XX
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9818931-A2.
XX
PD   07-MAY-1998.
XX
PF   30-OCT-1997;   97WO-US19588.
XX
PR   31-OCT-1996;   96US-0029960.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Choi GH,  Dillon PJ,  Dougherty BA,  Fannon M;
PI   Kunsch CA,  Rosen CA;
XX
DR   WPI; 1998-272225/24.
XX
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
XX
PS   Claim 1; Page 1381-1382; 1409pp; English.
XX
CC   The present invention describes a computer readable medium which has
CC   the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC   recorded on it, or a representative fragment or a sequence at least 95%
CC   identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC   SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC   Streptococcus pneumoniae. The present invention also describes an
CC   isolated nucleic acid molecule encoding a homologue of any of the
CC   fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC   nucleic acid molecule is produced by a process comprising: (a) screening
CC   a genomic DNA library using as a probe a target sequence defined by any
CC   of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC   library which contain sequences that hybridise to the target sequence and
CC   isolating the nucleic acid molecules from the members; or (b) isolating
CC   mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC   molecules whose nucleotide sequence is homologous to amplification
CC   primers derived from the fragment of the S. pneumoniae genome to prime
CC   the amplification and isolating the amplified sequences. The computer
CC   readable medium can be used in a computer-based system for identifying
CC   fragments of the S. pneumoniae genome of commercial importance, or
CC   expression modulating fragments of the S. pneumoniae genome. Products
CC   from the present invention can be used in diagnosis kits and assays, and
CC   pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ   Sequence 735 BP; 177 A; 191 C; 146 G; 220 T; 1 other;


  Query Match             64.8%;  Score 16.2;  DB 19;  Length 735;
  Best Local Similarity   85.7%;  Pred. No. 3e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagcacgttg 25
           | ||||| |||||||||| ||
Db      31 CGGGTGCTGGTCAGCACGGTG 11


RESULT  45
AAX30776
ID   AAX30776 standard; DNA; 1299 BP.
XX
AC   AAX30776;
XX
DT   20-MAY-1999  (first entry)
XX
DE   Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:53.
XX
KW   Streptococcus pneumoniae strain 0100993; vaccine; immune response;
KW   streptococcal infection; pneumococcal; ss.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9737026-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US05306.
XX
PR   22-AUG-1996;   96US-0025788.
PR   02-APR-1996;   96US-0014690.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Black MT,  Hodgson JE,  Knowles DJC,  Nicholas RO;
PI   Stodola RK;
XX
DR   WPI; 1997-503111/46.
DR   P-PSDB; AAY11177.
XX
PT   Nucleic acids encoding pneumococcal polypeptide(s) - useful in
PT   vaccines, drug screening, etc
XX
PS   Claim 5; Page 88-89; 354pp; English.
XX
CC   AAX30724 to AAX30946 represent genomic DNA sequences isolated from
CC   Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
CC   encode the novel proteins given in AAY11114 to AAY11367. The proteins,
CC   isolated from Streptococcus pneumoniae, can be used in vaccines against
CC   streptococcal infections and in assays for identifying compounds that
CC   inhibit or activate the activity of the proteins. The antagonists can
CC   be used to treat an individual having need to inhibit a bacterial
CC   protein. Vectors expressing the proteins can be used to induce a
CC   protective immune response in mammals.
XX
SQ   Sequence 1299 BP; 377 A; 261 C; 341 G; 320 T; 0 other;


  Query Match             64.8%;  Score 16.2;  DB 18;  Length 1299;
  Best Local Similarity   85.7%;  Pred. No. 3.2e+02;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagcacgttg 25
           | ||||| |||||||||| ||
Db    1030 cgggtgctggtcagcacggtg 1050


Search completed: December 26, 2001, 12:20:03
Job time: 6964 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:28:38 ; Search time 141.92 Seconds
(without alignments)
39.895 Million cell updates/sec

Title: US-09-396-196F-10
Perfect score: 25
Sequence: 1 tcgtcaggtgcaggtcagcacgttg 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
	2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
	3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
	4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
c	5	17.6	70.4	7286	5	PCT-US95-11684-1	Sequence 1, Appli
c	6	17	68.0	2095	1	US-08-336-408B-3	Sequence 3, Appli
c	7	17	68.0	2095	5	PCT-US91-00399-3	Sequence 3, Appli
c	8	17	68.0	20303	1	US-08-370-975B-6	Sequence 6, Appli
c	9	17	68.0	26764	1	US-08-370-975B-1	Sequence 1, Appli
	10	16.2	64.8	48	6	RE34606-11	Patent No. RE34,60
	11	16.2	64.8	49272	1	US-08-614-770A-1	Sequence 1, Appli
c	12	16	64.0	258	3	US-09-109-204-22	Sequence 22, Appl
c	13	16	64.0	1866	1	US-08-336-408B-1	Sequence 1, Appli
c	14	16	64.0	1866	5	PCT-US91-00399-1	Sequence 1, Appli
	15	16	64.0	2502	2	US-08-844-086-1	Sequence 1, Appli
	16	16	64.0	2502	3	US-09-018-211-1	Sequence 1, Appli
c	17	16	64.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	18	15.8	63.2	987	1	US-08-186-529-1	Sequence 1, Appli
	19	15.8	63.2	987	1	US-08-640-386A-1	Sequence 1, Appli
	20	15.8	63.2	987	4	US-08-848-760B-24	Sequence 24, Appl
	21	15.8	63.2	1018	2	US-08-184-009-194	Sequence 194, App
	22	15.8	63.2	1018	2	US-08-458-356-194	Sequence 194, App
	23	15.8	63.2	1018	4	US-08-460-736-194	Sequence 194, App
	24	15.8	63.2	1399	2	US-08-751-767A-3	Sequence 3, Appli
	25	15.8	63.2	1560	2	US-08-751-767A-11	Sequence 11, Appl
	26	15.8	63.2	1623	2	US-08-751-767A-9	Sequence 9, Appli
	27	15.8	63.2	2362	1	US-08-265-087-1	Sequence 1, Appli
	28	15.8	63.2	2362	1	US-08-621-493-1	Sequence 1, Appli
	29	15.8	63.2	2362	2	US-08-965-688-1	Sequence 1, Appli
	30	15.8	63.2	2362	4	US-09-260-173-1	Sequence 1, Appli
c	31	15.8	63.2	2655	1	US-08-471-033-17	Sequence 17, Appl
c	32	15.8	63.2	2655	1	US-08-471-033-26	Sequence 26, Appl
c	33	15.8	63.2	2655	2	US-08-471-044-17	Sequence 17, Appl
c	34	15.8	63.2	2655	2	US-08-471-044-26	Sequence 26, Appl
c	35	15.8	63.2	2655	2	US-08-463-483A-17	Sequence 17, Appl
c	36	15.8	63.2	2655	2	US-08-463-483A-26	Sequence 26, Appl
c	37	15.8	63.2	2655	2	US-08-471-046A-17	Sequence 17, Appl
c	38	15.8	63.2	2655	2	US-08-471-046A-26	Sequence 26, Appl
c	39	15.8	63.2	2655	2	US-08-470-566B-17	Sequence 17, Appl
c	40	15.8	63.2	2655	2	US-08-470-566B-26	Sequence 26, Appl
c	41	15.8	63.2	2655	2	US-08-469-334-17	Sequence 17, Appl
c	42	15.8	63.2	2655	2	US-08-469-334-26	Sequence 26, Appl
c	43	15.8	63.2	2655	3	US-09-300-529-17	Sequence 17, Appl
c	44	15.8	63.2	2655	3	US-09-300-529-26	Sequence 26, Appl
	45	15.8	63.2	6139	2	US-08-751-767A-7	Sequence 7, Appli
c	46	15.6	62.4	796	4	US-08-927-219-36	Sequence 36, Appl
c	47	15.6	62.4	907	2	US-08-486-663A-12	Sequence 12, Appl
c	48	15.6	62.4	907	3	US-08-767-942A-12	Sequence 12, Appl
c	49	15.6	62.4	907	5	PCT-US95-06722-23	Sequence 23, Appl
c	50	15.6	62.4	10014	4	US-08-927-219-130	Sequence 130, App
c	51	15.6	62.4	15894	1	US-08-348-891A-1	Sequence 1, Appli
c	52	15.6	62.4	15894	1	US-08-905-817-1	Sequence 1, Appli
c	53	15.6	62.4	31571	1	US-08-323-443B-1	Sequence 1, Appli
c	54	15.6	62.4	53526	3	US-08-658-136-2	Sequence 2, Appli
c	55	15.6	62.4	53577	3	US-08-658-136-1	Sequence 1, Appli
	56	15.6	62.4	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	57	15.6	62.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	58	15.4	61.6	630	3	US-08-545-809A-10	Sequence 10, Appl
c	59	15.4	61.6	1275	2	US-08-958-642-1	Sequence 1, Appli
c	60	15.4	61.6	1275	3	US-08-778-394-3	Sequence 3, Appli
c	61	15.4	61.6	1275	3	US-08-778-423A-1	Sequence 1, Appli
	62	15.2	60.8	604	4	US-09-328-111-391	Sequence 391, App
	63	15.2	60.8	632	4	US-09-328-111-798	Sequence 798, App
	64	15.2	60.8	699	4	US-08-998-416-678	Sequence 678, App
c	65	15.2	60.8	1235	1	US-08-457-797A-11	Sequence 11, Appl
c	66	15.2	60.8	1235	1	US-08-812-025-11	Sequence 11, Appl
c	67	15.2	60.8	1235	4	US-09-138-873A-11	Sequence 11, Appl
	68	15.2	60.8	1589	1	US-07-971-092-1	Sequence 1, Appli
	69	15.2	60.8	1611	6	5198342-1	Patent No. 5198342
c	70	15.2	60.8	2127	1	US-08-464-340A-1	Sequence 1, Appli
c	71	15.2	60.8	2127	5	PCT-US94-08449A-1	Sequence 1, Appli
	72	15.2	60.8	12412	1	US-08-390-878-18	Sequence 18, Appl
c	73	15.2	60.8	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	74	15	60.0	78	6	5164305-4	Patent No. 516430
	75	15	60.0	87	4	US-08-729-601A-16	Sequence 16, Appl
c	76	15	60.0	132	6	5164305-1	Patent No. 516430
c	77	15	60.0	437	2	US-08-721-746-3	Sequence 3, Appli
	78	15	60.0	475	4	US-08-991-789A-40	Sequence 40, Appl
c	79	15	60.0	477	4	US-08-991-789A-27	Sequence 27, Appl
c	80	15	60.0	1137	1	US-08-706-214-2	Sequence 2, Appli
	81	15	60.0	1265	4	US-08-991-789A-169	Sequence 169, App
	82	15	60.0	1591	2	US-08-646-981-14	Sequence 14, Appl
	83	15	60.0	1610	2	US-08-646-981-15	Sequence 15, Appl
c	84	15	60.0	1633	1	US-07-866-979-5	Sequence 5, Appli
c	85	15	60.0	1633	2	US-08-466-906B-5	Sequence 5, Appli
c	86	15	60.0	1633	3	US-08-706-281A-5	Sequence 5, Appli
c	87	15	60.0	1633	4	US-09-201-746-5	Sequence 5, Appli
c	88	15	60.0	1633	4	US-09-097-231-5	Sequence 5, Appli
	89	15	60.0	1646	4	US-08-991-789A-290	Sequence 290, App
c	90	15	60.0	1815	4	US-09-178-252-3	Sequence 3, Appli
	91	15	60.0	1833	1	US-08-454-720A-36	Sequence 36, Appl
c	92	15	60.0	1840	1	US-08-399-696-126	Sequence 126, App
c	93	15	60.0	1849	1	US-08-399-696-3	Sequence 3, Appli
c	94	15	60.0	1856	1	US-08-706-214-1	Sequence 1, Appli
c	95	15	60.0	1962	1	US-08-399-696-1	Sequence 1, Appli
c	96	15	60.0	2520	4	US-08-931-608A-2	Sequence 2, Appli
c	97	15	60.0	2600	3	US-08-957-063-1	Sequence 1, Appli
	98	15	60.0	3273	1	US-08-148-122A-2	Sequence 2, Appli
c	99	15	60.0	3393	1	US-08-295-502-1	Sequence 1, Appli
c	100	15	60.0	3393	5	PCT-US95-10691-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Enhanced Biotin Biosynthesis in Plant Tissue
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/351,970
;      FILING DATE:  08-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-401-068-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.017;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    111 TCGTCAGGTGCAGGTCAGCACGTTG 135


RESULT   2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-846-338-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.017;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    111 TCGTCAGGTGCAGGTCAGCACGTTG 135


RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A-15/9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1157
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 117
;      OTHER INFORMATION:  /product= "Biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioB"
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2295..3050
;      OTHER INFORMATION:  /codon_start= 2295
;      OTHER INFORMATION:  /function= "involved in pimeloyl-CoA synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /gene= "bioC"
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3750..5039
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3750
;      OTHER INFORMATION:  /EC_number= 2.6.1.62
;      OTHER INFORMATION:  /product= "DAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioA"
;      OTHER INFORMATION:  /number= 5
;      OTHER INFORMATION:  /standard_name=
;      OTHER INFORMATION:  "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
;      OTHER INFORMATION:  aminotransf."
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  5098..5574
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 5098
;      OTHER INFORMATION:  /function= "unknown, involved in biotin synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "ORFI"
;      OTHER INFORMATION:  /number= 6
;    FEATURE:
;      NAME/KEY:  -10_signal
;      LOCATION:  45..49
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  -35_signal
;      LOCATION:  23..28
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  105..119
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "bioB RBS no.9"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  2284..2297
;      OTHER INFORMATION:  /standard_name= "bioC RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3742..3752
;      OTHER INFORMATION:  /standard_name= "bioA RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  5088..5100
;      OTHER INFORMATION:  /standard_name= "ORFI RBS"
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  5583..5644
;      OTHER INFORMATION:  /standard_name= "rho-independent
;      OTHER INFORMATION:  transcriptional terminator"
;    FEATURE:
;      NAME/KEY:  stem_loop
;      LOCATION:  5583..5605
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..96
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "promoter ptac"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-1


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.02;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT   4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A15-9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1154..2308
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1154
;      OTHER INFORMATION:  /EC_number= 2.3.1.47
;      OTHER INFORMATION:  /product= "KAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "8-Amino-7-oxononanoate synthase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3043..3753
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3043
;      OTHER INFORMATION:  /EC_number= 6.3.3.3
;      OTHER INFORMATION:  /product= "DTB synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioD"
;      OTHER INFORMATION:  /number= 4
;      OTHER INFORMATION:  /standard_name= "Dethiobiotin synthase"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1141..1156
;      OTHER INFORMATION:  /standard_name= "bioF RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3030..3045
;      OTHER INFORMATION:  /standard_name= "bioD RBS"
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-6


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.02;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT   5
PCT-US95-11684-1/c
; Sequence 1, Application PC/TUS9511684
;  GENERAL INFORMATION:
;    APPLICANT:  THE SCRIPPS RESEARCH INSTITUTE
;    TITLE OF INVENTION:  CYTOTACTIN DERIVATIVES THAT STIMULATE
;    TITLE OF INVENTION:  ATTACHMENT AND NEURITE OUTGROWTH, AND METHODS OF MAKING
;    TITLE OF INVENTION:  AND USING SAME
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute, Office of
;      ADDRESSEE:  Patent Counsel
;      STREET:  10666 North Torrey Pines Road, TPC 8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/11684
;      FILING DATE:  14-SEP-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,359
;      FILING DATE:  16-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Logan, April C.
;      REGISTRATION NUMBER:  33,950
;      REFERENCE/DOCKET NUMBER:  BEC0019P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-554-2937
;      TELEFAX:  619-554-6312
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7286 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  55..6654
;      OTHER INFORMATION:  /product= "cytotactin"
PCT-US95-11684-1


  Query Match             70.4%;  Score 17.6;  DB 5;  Length 7286;
  Best Local Similarity   83.3%;  Pred. No. 33;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacgttg 25
          | ||||||||||||| || |||| |
Db   2200 CTTCAGGTGCAGGTAAGTACGTGG 2177


RESULT   6
US-08-336-408B-3/c
; Sequence 3, Application US/08336408B
; Patent No. 5723329
;  GENERAL INFORMATION:
;    APPLICANT:  EVANS, RONALD M.
;    APPLICANT:  MANGELSDORF, DAVID J.
;    TITLE OF INVENTION:  RETINOID RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
;      STREET:  444 SOUTH FLOWER STREET, SUITE 2000
;      CITY:  LOS ANGELES
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/336,408B
;      FILING DATE:  08-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/933,453
;      FILING DATE:  21-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US91/00399
;      FILING DATE:  22-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/478,071
;      FILING DATE:  09-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  REITER, STEPHEN E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9851
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-1995
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2095 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  178..1581
US-08-336-408B-3


  Query Match             68.0%;  Score 17;  DB 1;  Length 2095;
  Best Local Similarity   80.0%;  Pred. No. 53;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           ||| |||||| ||||||| | |||
Db     708 TCGGCAGGTGTAGGTCAGGTCTTTG 684


RESULT   7
PCT-US91-00399-3/c
; Sequence 3, Application PC/TUS9100399
;  GENERAL INFORMATION:
;    APPLICANT:  Mangelsdorf, Dr., David J.
;    APPLICANT:  Evans Dr., Ronald M.
;    TITLE OF INVENTION:  RETINOID RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FITCH, EVEN, TABIN & FLANNERY
;      STREET:  135 South LaSalle Street, Suite 900
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/00399
;      FILING DATE:  19910122
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/478,071
;      FILING DATE:  09-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Mr., Stephen E.
;      REGISTRATION NUMBER:  31192
;      REFERENCE/DOCKET NUMBER:  50852
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 552-1311
;      TELEFAX:  (619) 552-0095
;      TELEX:  20 6566 PATLAW CGO
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2095 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  RXR MOUSE ALPHA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(178..1581)
PCT-US91-00399-3


  Query Match             68.0%;  Score 17;  DB 5;  Length 2095;
  Best Local Similarity   80.0%;  Pred. No. 53;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           ||| |||||| ||||||| | |||
Db     708 TCGGCAGGTGTAGGTCAGGTCTTTG 684


RESULT   8
US-08-370-975B-6/c
; Sequence 6, Application US/08370975B
; Patent No. 5622851
;  GENERAL INFORMATION:
;    APPLICANT:  Maley, Frank
;    APPLICANT:  Maley, Gladys F.
;    APPLICANT:  Weiner, Karen X.B.
;    TITLE OF INVENTION:  Human Deoxycytidylate Deaminase Gene
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/370,975B
;      FILING DATE:  10-JAN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Timian, Susan J.
;      REGISTRATION NUMBER:  34,103
;      REFERENCE/DOCKET NUMBER:  20894/80
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716)263-1636
;      TELEFAX:  (716)263-1600
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20303 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
```

```
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  4q35
US-08-370-975B-6


  Query Match             68.0%;  Score 17;  DB 1;  Length 20303;
  Best Local Similarity   80.0%;  Pred. No. 64;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          | ||  |||||||| ||||| ||||
Db    285 TGGTGTGGTGCAGGCCAGCAGGTTG 261


RESULT   9
US-08-370-975B-1/c
; Sequence 1, Application US/08370975B
; Patent No. 5622851
;  GENERAL INFORMATION:
;    APPLICANT:  Maley, Frank
;    APPLICANT:  Maley, Gladys F.
;    APPLICANT:  Weiner, Karen X.B.
;    TITLE OF INVENTION:  Human Deoxycytidylate Deaminase Gene
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/370,975B
;      FILING DATE:  10-JAN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Timian, Susan J.
;      REGISTRATION NUMBER:  34,103
;      REFERENCE/DOCKET NUMBER:  20894/80
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716)263-1636
;      TELEFAX:  (716)263-1600
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  26764 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  4q35
US-08-370-975B-1


  Query Match             68.0%;  Score 17;  DB 1;  Length 26764;
  Best Local Similarity   80.0%;  Pred. No. 66;
  Matches   20;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          | ||  |||||||| ||||| ||||
Db   2248 TGGTGTGGTGCAGGCCAGCAGGTTG 2224


RESULT  10
RE34606-11
;Patent No. RE34,606
;    APPLICANT: ESTELL, DAVID A.;WELLS, JAMES A.;BOTT,
;RICHARD R.
;    TITLE OF INVENTION: MODIFIED ENZYMES AND METHODS FOR
;MAKING SAME
;    NUMBER OF SEQUENCES: 27
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/556,918
;      FILING DATE: 20-JUL-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 614,612
;      FILING DATE: 29-MAY-1984
;SEQ ID NO:11:
;      LENGTH: 48
RE34606-11


  Query Match             64.8%;  Score 16.2;  DB 6;  Length 48;
  Best Local Similarity   75.0%;  Pred. No. 87;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacgttg 25
          |||||   |||  || ||||||||
Db     14 cgtcannngcatctccgcacgttg 37


RESULT  11
US-08-614-770A-1
; Sequence 1, Application US/08614770A
; Patent No. 5773267
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAM R. JACOBS AND GRAHAM F. HATFULL
;    TITLE OF INVENTION:  D29 SHUTTLE PHASMIDS AND USES THEREOF
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  AMSTER, ROTHSTEIN & EBENSTEIN
;      STREET:  90 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  U.S.A.
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH 1.44 Mb STORAGE
;      MEDIUM TYPE:  DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/614,770A
;      FILING DATE:  MARCH 7, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ELIZABETH A. BOGOSIAN
;      REGISTRATION NUMBER:  39,911
;      REFERENCE/DOCKET NUMBER:  96700/402
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  49272
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:
;      DESCRIPTION:  OLIGONUCLEOTIDE
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  MYCOBACTERIOPHAGE
;      INDIVIDUAL ISOLATE:  D29
US-08-614-770A-1


  Query Match             64.8%;  Score 16.2;  DB 1;  Length 49272;
  Best Local Similarity   85.7%;  Pred. No. 1.5e+02;
```

```
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcac 21
          || ||||||| ||||| ||||
Db  30020 TCATCAGGTGGAGGTCCGCAC 30040


RESULT  12
US-09-109-204-22/c
; Sequence 22, Application US/09109204
; Patent No. 6060250
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Gorgone, Gina A.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Patterson, Chandra
;    TITLE OF INVENTION:  HUMAN TRANSFERASES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/109,204
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cerrone, Michael C
;      REGISTRATION NUMBER:  39,132
;      REFERENCE/DOCKET NUMBER:  PF-0546 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-855-0572
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  258 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  BLADTUT07
;      CLONE:  1889292H1
US-09-109-204-22


  Query Match             64.0%;  Score 16;  DB 3;  Length 258;
  Best Local Similarity   76.0%;  Pred. No. 1.2e+02;
  Matches   19;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          | || |||||| ||| |||||| |
Db    257 TGGTAAGGTGCNTGTCCGCACGTAG 233


RESULT  13
US-08-336-408B-1/c
; Sequence 1, Application US/08336408B
; Patent No. 5723329
;  GENERAL INFORMATION:
;    APPLICANT:  EVANS, RONALD M.
;    APPLICANT:  MANGELSDORF, DAVID J.
;    TITLE OF INVENTION:  RETINOID RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
;      STREET:  444 SOUTH FLOWER STREET, SUITE 2000
;      CITY:  LOS ANGELES
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/336,408B
;      FILING DATE:  08-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/933,453
;      FILING DATE:  21-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US91/00399
;      FILING DATE:  22-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/478,071
;      FILING DATE:  09-FEB-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  REITER, STEPHEN E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9851
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-1995
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1866 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  76..1464
US-08-336-408B-1


  Query Match             64.0%;  Score 16;  DB 1;  Length 1866;
  Best Local Similarity   79.2%;  Pred. No. 1.4e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacgttg 25
          || |||||| ||||||| | |||
Db    590 CGGCAGGTGTAGGTCAGGTCCTTG 567


RESULT  14
PCT-US91-00399-1/c
; Sequence 1, Application PC/TUS9100399
;  GENERAL INFORMATION:
;    APPLICANT:  Mangelsdorf, Dr., David J.
;    APPLICANT:  Evans Dr., Ronald M.
;    TITLE OF INVENTION:  RETINOID RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FITCH, EVEN, TABIN & FLANNERY
;      STREET:  135 South LaSalle Street, Suite 900
;      CITY:  Chicago
```

```
;       STATE:  Illinois
;       COUNTRY:  USA
;       ZIP:  60603
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US91/00399
;       FILING DATE:  19910122
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/478,071
;       FILING DATE:  09-FEB-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Reiter, Mr., Stephen E.
;       REGISTRATION NUMBER:  31192
;       REFERENCE/DOCKET NUMBER:  50852
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (619) 552-1311
;       TELEFAX:  (619) 552-0095
;       TELEX:  20 6566 PATLAW CGO
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1866 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     IMMEDIATE SOURCE:
;       CLONE:  RXR HUMAN ALPHA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  join(76..1464)
PCT-US91-00399-1


  Query Match             64.0%;  Score 16;  DB 5;  Length 1866;
  Best Local Similarity   79.2%;  Pred. No. 1.4e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           || ||||||| ||||||| | |||
Db     590 CGGCAGGTGTAGGTCAGGTCCTTG 567


RESULT  15
US-08-844-086-1
; Sequence 1, Application US/08844086
; Patent No. 5866390
;  GENERAL INFORMATION:
;    APPLICANT:  Lawlor, Elizabeth
;    TITLE OF INVENTION:  No. 5866390el Compounds
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/844,086
;      FILING DATE:  18-APR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  9607993.4
;      FILING DATE:  18-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P31457-4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2502 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-844-086-1


  Query Match             64.0%;  Score 16;  DB 2;  Length 2502;
  Best Local Similarity   79.2%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           |||| || || |||| ||||||| |
Db     128 CGTCTGGAGCTGGTCTGCACGTAG 151


RESULT  16
US-09-018-211-1
; Sequence 1, Application US/09018211
; Patent No. 6048716
;  GENERAL INFORMATION:
;    APPLICANT:  Lawlor, Elizabeth
;    TITLE OF INVENTION:  No. 6048716el Compounds
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/018,211
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/844,086
;      FILING DATE:  18-APR-1997
;      APPLICATION NUMBER:  9607993.4
;      FILING DATE:  18-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gimmi, Edward R
;      REGISTRATION NUMBER:  38,891
;      REFERENCE/DOCKET NUMBER:  P31457-4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-4478
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2502 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-09-018-211-1
```

```
  Query Match             64.0%;  Score 16;  DB 3;  Length 2502;
  Best Local Similarity   79.2%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacgttg 25
          |||| || || |||| ||||||  |
Db    128 CGTCTGGAGCTGGTCTGCACGTAG 151


RESULT  17
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


  Query Match             64.0%;  Score 16;  DB 4;  Length 4403765;
  Best Local Similarity   83.3%;  Pred. No. 51;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacgttg 25
          |||| || |  ||||||||||||
Db 752331 CGTCCGGCGTGGGTCAGCACGTTG 752308


RESULT  18
US-08-186-529-1
; Sequence 1, Application US/08186529
; Patent No. 5573764
;  GENERAL INFORMATION:
;    APPLICANT:  Sykes, Megan
;    APPLICANT:  Wolf, Stanley F.
;    TITLE OF INVENTION:  USE OF INTERLEUKIN-12 TO PREVENT
;    TITLE OF INVENTION:  GRAFT-VERSUS-HOST DISEASE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc., Legal Affairs
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/186,529
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McDaniels, Patricia A.
;      REGISTRATION NUMBER:  33,194
;      REFERENCE/DOCKET NUMBER:  GI 5225
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-498-8401
;      TELEFAX:  617-876-5851
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  987 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL TYPE:  Lymphoblast
;      CELL LINE:  RPMI 8866
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..987
US-08-186-529-1


  Query Match             63.2%;  Score 15.8;  DB 1;  Length 987;
  Best Local Similarity   89.5%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagc 19
          ||| |||||| ||||||||
Db    747 TCGGCAGGTGGAGGTCAGC 765


RESULT  19
US-08-640-386A-1
; Sequence 1, Application US/08640386A
; Patent No. 5756085
;  GENERAL INFORMATION:
;    APPLICANT:  Sykes, Megan
;    APPLICANT:  Wolf, Stanley F.
;    TITLE OF INVENTION:  USE OF INTERLEUKIN-12 TO PREVENT
;    TITLE OF INVENTION:  GRAFT-VERSUS-HOST DISEASE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc., Legal Affairs
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/640,386A
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI 5225A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-498-8224
;      TELEFAX:  617-876-5851
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  987 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
```

```
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL TYPE:  Lymphoblast
;      CELL LINE:  RPMI 8866
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..987
US-08-640-386A-1


  Query Match             63.2%;  Score 15.8;  DB 1;  Length 987;
  Best Local Similarity   89.5%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagc 19
          ||| |||||| ||||||||
Db    747 TCGGCAGGTGGAGGTCAGC 765


RESULT  20
US-08-848-760B-24
; Sequence 24, Application US/08848760B
; Patent No. 6248721
;  GENERAL INFORMATION:
;       APPLICANT: Chang, Lung-Ji
;       TITLE OF INVENTION: Animal Model For Evaluation Of Vaccines
;       NUMBER OF SEQUENCES: 33
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
;            STREET: 2421 N.W. 41st Street, Suite A-1
;            CITY: Gainesville
;            STATE: Florida
;            COUNTRY: United States of America
;            ZIP: 32606
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/848,760B
;            FILING DATE: 25-Jan-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/838,702
;            FILING DATE: 09-APR-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: PACE, DORAN R.
;            REGISTRATION NUMBER: 38,261
;            REFERENCE/DOCKET NUMBER: CNG-100C1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (352) 375-8100
;            TELEFAX: (352) 372-5800
;  INFORMATION FOR SEQ ID NO: 24:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 987 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: other nucleic acid
;            DESCRIPTION: /desc = "DNA"
;       SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-08-848-760B-24


  Query Match             63.2%;  Score 15.8;  DB 4;  Length 987;
  Best Local Similarity   89.5%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagc 19
          ||| |||||| ||||||||
Db    747 TCGGCAGGTGGAGGTCAGC 765


RESULT  21
US-08-184-009-194
; Sequence 194, Application US/08184009
; Patent No. 5833975
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;    NUMBER OF SEQUENCES:  217
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/184,009
;      FILING DATE:  19-JAN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2530
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;      TELEX:  425066CURTMS
;  INFORMATION FOR SEQ ID NO:  194:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1018 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-184-009-194


  Query Match             63.2%;  Score 15.8;  DB 2;  Length 1018;
  Best Local Similarity   89.5%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagc 19
          ||| |||||| ||||||||
Db    778 TCGGCAGGTGGAGGTCAGC 796


RESULT  22
US-08-458-356-194
; Sequence 194, Application US/08458356
; Patent No. 5942235
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    APPLICANT:  Tartaglia, James
;    APPLICANT:  Cox, William I.
;    TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;    NUMBER OF SEQUENCES:  217
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/458,356
;       FILING DATE:  02-JUN-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/184,009
;       FILING DATE:  19-JAN-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Frommer, William S.
;       REGISTRATION NUMBER:  25,506
;       REFERENCE/DOCKET NUMBER:  454310-2530
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 840-3333
;       TELEFAX:  (212) 840-0712
;       TELEX:  425066CURTMS
;   INFORMATION FOR SEQ ID NO:  194:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1018 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-458-356-194


  Query Match             63.2%;  Score 15.8;  DB 2;  Length 1018;
  Best Local Similarity   89.5%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagc 19
          ||| |||||| ||||||||
Db    778 TCGGCAGGTGGAGGTCAGC 796


RESULT  23
US-08-460-736-194
; Sequence 194, Application US/08460736
; Patent No. 6265189
;  GENERAL INFORMATION:
;     APPLICANT:  Paoletti, Enzo
;     APPLICANT:  Tartaglia, James
;     APPLICANT:  Cox, William I.
;     TITLE OF INVENTION:  RECOMBINANT VIRUS IMMUNOTHERAPY
;     NUMBER OF SEQUENCES:  217
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Curtis, Morris & Safford
;       STREET:  530 Fifth Avenue
;       CITY:  New York
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/460,736
;       FILING DATE:  02-JUN-1995
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/184,009
;       FILING DATE:  19-JAN-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Frommer, William S.
;       REGISTRATION NUMBER:  25,506
;       REFERENCE/DOCKET NUMBER:  454310-2530
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 840-3333
;       TELEFAX:  (212) 840-0712
;       TELEX:  425066CURTMS
;   INFORMATION FOR SEQ ID NO:  194:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1018 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-460-736-194


  Query Match             63.2%;  Score 15.8;  DB 4;  Length 1018;
  Best Local Similarity   89.5%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagc 19
          ||| |||||| ||||||||
Db    778 TCGGCAGGTGGAGGTCAGC 796


RESULT  24
US-08-751-767A-3
; Sequence 3, Application US/08751767A
; Patent No. 5994104
;  GENERAL INFORMATION:
;     APPLICANT:  ANDERSON, ROBERT J.
;     APPLICANT:  GRANT, HUGH
;     APPLICANT:  MACDONALD, IAN D.
;     TITLE OF INVENTION:  INTERLUKIN-12 FUSION PROTEIN
;     NUMBER OF SEQUENCES:  80
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  NIXON & VANDERHYE P.C.
;       STREET:  1100 NORTH GLEBE ROAD
;       CITY:  ARLINGTON
;       STATE:  VA
;       COUNTRY:  USA
;       ZIP:  22201
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/751,767A
;       FILING DATE:  08-NOV-1996
;       CLASSIFICATION:  536
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  SADOFF, B.J.
;       REGISTRATION NUMBER:  36,663
;       REFERENCE/DOCKET NUMBER:  117-221
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  7038164091
;       TELEFAX:  7038164100
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1399 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  43..1026
US-08-751-767A-3


  Query Match             63.2%;  Score 15.8;  DB 2;  Length 1399;
  Best Local Similarity   89.5%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagc 19
          ||| |||||| ||||||||
Db    789 TCGGCAGGTGGAGGTCAGC 807
```

```
RESULT  25
US-08-751-767A-11
; Sequence 11, Application US/08751767A
; Patent No. 5994104
;  GENERAL INFORMATION:
;    APPLICANT:  ANDERSON, ROBERT J.
;    APPLICANT:  GRANT, HUGH
;    APPLICANT:  MACDONALD, IAN D.
;    TITLE OF INVENTION:  INTERLUKIN-12 FUSION PROTEIN
;    NUMBER OF SEQUENCES:  80
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/751,767A
;      FILING DATE:  08-NOV-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SADOFF, B.J.
;      REGISTRATION NUMBER:  36,663
;      REFERENCE/DOCKET NUMBER:  117-221
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  7038164091
;      TELEFAX:  7038164100
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1557
US-08-751-767A-11


  Query Match             63.2%;  Score 15.8;  DB 2;  Length 1560;
  Best Local Similarity   89.5%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagc 19
           ||| |||||| ||||||||
Db     684 TCGGCAGGTGGAGGTCAGC 702


RESULT  26
US-08-751-767A-9
; Sequence 9, Application US/08751767A
; Patent No. 5994104
;  GENERAL INFORMATION:
;    APPLICANT:  ANDERSON, ROBERT J.
;    APPLICANT:  GRANT, HUGH
;    APPLICANT:  MACDONALD, IAN D.
;    TITLE OF INVENTION:  INTERLUKIN-12 FUSION PROTEIN
;    NUMBER OF SEQUENCES:  80
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/751,767A
;      FILING DATE:  08-NOV-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SADOFF, B.J.
;      REGISTRATION NUMBER:  36,663
;      REFERENCE/DOCKET NUMBER:  117-221
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  7038164091
;      TELEFAX:  7038164100
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1623 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1620
US-08-751-767A-9


  Query Match             63.2%;  Score 15.8;  DB 2;  Length 1623;
  Best Local Similarity   89.5%;  Pred. No. 1.7e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagc 19
           ||| |||||| ||||||||
Db     747 TCGGCAGGTGGAGGTCAGC 765


RESULT  27
US-08-265-087-1
; Sequence 1, Application US/08265087
; Patent No. 5571515
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, Phillip
;    APPLICANT:  Trinchieri, Giorgio
;    TITLE OF INVENTION:  Compositions and Methods for Use of
;    TITLE OF INVENTION:  IL-12 as an Adjuvant
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Center, PO Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/265,087
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/229,282
;      FILING DATE:  18-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  WST51AUSA
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2362 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1016
US-08-265-087-1


  Query Match             63.2%;  Score 15.8;  DB 1;  Length 2362;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagc 19
          ||| |||||| ||||||||
Db    779 TCGGCAGGTGGAGGTCAGC 797


RESULT  28
US-08-621-493-1
; Sequence 1, Application US/08621493
; Patent No. 5723127
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, Phillip
;    APPLICANT:  Trinchieri, Giorgio
;    TITLE OF INVENTION:  Compositions and Methods for Use of
;    TITLE OF INVENTION:  IL-12 as an Adjuvant
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Center, PO Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/621,493
;      FILING DATE:  25-MAR-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/265,087
;      FILING DATE:  17-JUN-1994
;      APPLICATION NUMBER:  US 08/229,282
;      FILING DATE:  18-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  WST51AUSA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2362 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1016
US-08-621-493-1
```

```
  Query Match             63.2%;  Score 15.8;  DB 1;  Length 2362;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagc 19
          ||| |||||| ||||||||
Db    779 TCGGCAGGTGGAGGTCAGC 797


RESULT  29
US-08-965-688-1
; Sequence 1, Application US/08965688
; Patent No. 5976539
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, Phillip
;    APPLICANT:  Trinchieri, Giorgio
;    TITLE OF INVENTION:  Compositions and Methods for Use of
;    TITLE OF INVENTION:  IL-12 as an Adjuvant
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Center, PO Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/965,688
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/621,493
;      FILING DATE:  25-MAR-1996
;      APPLICATION NUMBER:  08/265,087
;      FILING DATE:  17-JUN-1994
;      APPLICATION NUMBER:  US 08/229,282
;      FILING DATE:  18-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  WST51AUSA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2362 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  33..1016
US-08-965-688-1


  Query Match             63.2%;  Score 15.8;  DB 2;  Length 2362;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagc 19
          ||| |||||| ||||||||
Db    779 TCGGCAGGTGGAGGTCAGC 797
```

```
RESULT  30
US-09-260-173-1
; Sequence 1, Application US/09260173
; Patent No. 6168923
;   GENERAL INFORMATION:
;        APPLICANT: Scott, Phillip
;                   Trinchieri, Giorgio
;        TITLE OF INVENTION: Compositions and Methods for Use of
;                            IL-12 as an Adjuvant
;        NUMBER OF SEQUENCES: 4
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Howson and Howson
;             STREET: Spring House Corporate Center, PO Box 457
;             CITY: Spring House
;             STATE: Pennsylvania
;             COUNTRY: USA
;             ZIP: 19477
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/260,173
;             FILING DATE: 01-Mar-1999
;             CLASSIFICATION: <Unknown>
;                             17-JUN-1994
;                             18-APR-1994
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/965,688
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: 08/265,087
;             FILING DATE: 17-JUN-1994
;             APPLICATION NUMBER: US 08/229,282
;             FILING DATE: 18-APR-1994
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Bak, Mary E.
;             REGISTRATION NUMBER: 31,215
;             REFERENCE/DOCKET NUMBER: WST51AUSA
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 215-540-9206
;             TELEFAX: 215-540-5818
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 2362 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: unknown
;        MOLECULE TYPE: cDNA
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  33..1016
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-260-173-1


  Query Match             63.2%;  Score 15.8;  DB 4;  Length 2362;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagc 19
           ||| |||||| ||||||||
Db     779 TCGGCAGGTGGAGGTCAGC 797


RESULT  31
US-08-471-033-17/c
; Sequence 17, Application US/08471033
; Patent No. 5770696
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5770696el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,033
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  P-40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV7 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2652
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence for 100 kd VIP1A(a) protein from AB78"
US-08-471-033-17


  Query Match             63.2%;  Score 15.8;  DB 1;  Length 2655;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagcacgttg 25
           |||||||||| |||| ||
Db      45 GGTGCAGGTCACCACGCTG 27


RESULT  32
US-08-471-033-26/c
; Sequence 26, Application US/08471033
; Patent No. 5770696
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
```

```
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5770696el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,033
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  P-40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV7 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2655
;      OTHER INFORMATION:  /note= "maize optimized DNA
;      OTHER INFORMATION:  sequence encoding VIP1A(a)"
US-08-471-033-26


  Query Match             63.2%;  Score 15.8;  DB 1;  Length 2655;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagcacgttg 25
           |||||||||| |||| ||
Db      45 GGTGCAGGTCACCACGCTG 27


RESULT  33
US-08-471-044-17/c
; Sequence 17, Application US/08471044
; Patent No. 5840868
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5840868el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,044
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV6 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2652
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence for 100 kd VIP1A(a) protein from AB78"
US-08-471-044-17


  Query Match             63.2%;  Score 15.8;  DB 2;  Length 2655;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagcacgttg 25
           |||||||||| |||| ||
Db      45 GGTGCAGGTCACCACGCTG 27


RESULT  34
US-08-471-044-26/c
```

```
; Sequence 26, Application US/08471044
; Patent No. 5840868
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5840868el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,044
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pace, Gary M.
;      REGISTRATION NUMBER:  40,403
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3/DIV6 - SQLv3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8582
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2655
;      OTHER INFORMATION:  /note= "maize optimized DNA
;      OTHER INFORMATION:  sequence encoding VIP1A(a)"
US-08-471-044-26


  Query Match           63.2%;  Score 15.8;  DB 2;  Length 2655;
  Best Local Similarity  89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagcacgttg 25
          ||||||||||| |||| ||
Db     45 GGTGCAGGTCACCACGCTG 27
```

```
RESULT  35
US-08-463-483A-17/c
; Sequence 17, Application US/08463483A
; Patent No. 5849870
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5849870el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,483A
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2652
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence for 100 kd VIP1A(a) protein from AB78"
US-08-463-483A-17


  Query Match           63.2%;  Score 15.8;  DB 2;  Length 2655;
  Best Local Similarity  89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagcacgttg 25
          ||||||||||| |||| ||
```

```
Db      45 GGTGCAGGTCACCACGCTG 27


RESULT  36
US-08-463-483A-26/c
; Sequence 26, Application US/08463483A
; Patent No. 5849870
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5849870el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,483A
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2655
;      OTHER INFORMATION:  /note= "maize optimized DNA
;      OTHER INFORMATION:  sequence encoding VIP1A(a)"
US-08-463-483A-26


  Query Match             63.2%;  Score 15.8;  DB 2;  Length 2655;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagcacgttg 25
                 |||||||||| |||| ||
Db      45 GGTGCAGGTCACCACGCTG 27


RESULT  37
US-08-471-046A-17/c
; Sequence 17, Application US/08471046A
; Patent No. 5866326
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  Method For Isolating Vegetative Insecticidal
;    TITLE OF INVENTION:  Protein Genes
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5866326artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,046A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1695/CIP3/DIV8 - SQLv4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2652
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence for 100 kd VIP1A(a) protein from AB78"
US-08-471-046A-17
```

```
  Query Match             63.2%;  Score 15.8;  DB 2;  Length 2655;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagcacgttg 25
           |||||||||||| |||| ||
Db      45 GGTGCAGGTCACCACGCTG 27


RESULT  38
US-08-471-046A-26/c
; Sequence 26, Application US/08471046A
; Patent No. 5866326
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  Method For Isolating Vegetative Insecticidal
;    TITLE OF INVENTION:  Protein Genes
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5866326artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,046A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1695/CIP3/DIV8 - SQLv4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2655
;      OTHER INFORMATION:  /note= "maize optimized DNA
;      OTHER INFORMATION:  sequence encoding VIP1A(a)"
US-08-471-046A-26


  Query Match             63.2%;  Score 15.8;  DB 2;  Length 2655;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagcacgttg 25
           |||||||||||| |||| ||
Db      45 GGTGCAGGTCACCACGCTG 27


RESULT  39
US-08-470-566B-17/c
; Sequence 17, Application US/08470566B
; Patent No. 5872212
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5872212el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5872212artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,566B
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1695/CIP3/DIV4 - SQLv4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
```

```
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2652
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence for 100 kd VIP1A(a) protein from AB78"
US-08-470-566B-17


  Query Match             63.2%;  Score 15.8;  DB 2;  Length 2655;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagcacgttg 25
          |||||||||||  |||| ||
Db     45 GGTGCAGGTCACCACGCTG 27


RESULT  40
US-08-470-566B-26/c
; Sequence 26, Application US/08470566B
; Patent No. 5872212
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5872212el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5872212artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,566B
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC1695/CIP3/DIV4 - SQLv4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2655
;      OTHER INFORMATION:  /note= "maize optimized DNA
;      OTHER INFORMATION:  sequence encoding VIP1A(a)"
US-08-470-566B-26


  Query Match             63.2%;  Score 15.8;  DB 2;  Length 2655;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagcacgttg 25
          |||||||||||  |||| ||
Db     45 GGTGCAGGTCACCACGCTG 27


RESULT  41
US-08-469-334-17/c
; Sequence 17, Application US/08469334
; Patent No. 5990383
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5990383el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,334
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/463,483
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1695/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  17:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2652
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence for 100 kd VIP1A(a) protein from AB78"
US-08-469-334-17


  Query Match             63.2%;  Score 15.8;  DB 2;  Length 2655;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagcacgttg 25
           ||||||||||| |||| ||
Db      45 GGTGCAGGTCACCACGCTG 27


RESULT  42
US-08-469-334-26/c
; Sequence 26, Application US/08469334
; Patent No. 5990383
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  No. 5990383el Pesticidal Proteins and Strains
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,334
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/463,483
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1695/C1P3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2655
;      OTHER INFORMATION:  /note= "maize optimized DNA
;      OTHER INFORMATION:  sequence encoding VIP1A(a)"
US-08-469-334-26


  Query Match             63.2%;  Score 15.8;  DB 2;  Length 2655;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagcacgttg 25
           ||||||||||| |||| ||
Db      45 GGTGCAGGTCACCACGCTG 27


RESULT  43
US-09-300-529-17/c
; Sequence 17, Application US/09300529
; Patent No. 6066783
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  Genes Encoding Insecticidal Proteins
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6066783artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/300,529
;      FILING DATE:  TBA
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/469,334
;      FILING DATE:  06-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  S-19506L
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2652
;      OTHER INFORMATION:  /note= "Maize optimized DNA
;      OTHER INFORMATION:  sequence for 100 kd VIP1A(a) protein from AB78"
US-09-300-529-17


  Query Match             63.2%;  Score 15.8;  DB 3;  Length 2655;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagcacgttg 25
           |||||||||||  |||| ||
Db      45 GGTGCAGGTCACCACGCTG 27


RESULT  44
US-09-300-529-26/c
; Sequence 26, Application US/09300529
; Patent No. 6066783
;  GENERAL INFORMATION:
;    APPLICANT:  Warren, Gregory W
;    APPLICANT:  Koziel, Michael G
;    APPLICANT:  Mullins, Martha A
;    APPLICANT:  Nye, Gordon J
;    APPLICANT:  Carr, Brian
;    APPLICANT:  Desai, Nalini M
;    APPLICANT:  Kostichka, N. Kristy
;    APPLICANT:  Duck, Nicholas B
;    APPLICANT:  Estruch, Juan J
;    TITLE OF INVENTION:  Genes Encoding Insecticidal Proteins
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6066783artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/300,529
;      FILING DATE:  TBA
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/469,334
;      FILING DATE:  06-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/463,483
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/314,594
;      FILING DATE:  09-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/218,018
;      FILING DATE:  23-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/037,057
;      FILING DATE:  25-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  S-19506L
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "Synthetic DNA"
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2655
;      OTHER INFORMATION:  /note= "maize optimized DNA
;      OTHER INFORMATION:  sequence encoding VIP1A(a)"
US-09-300-529-26


  Query Match             63.2%;  Score 15.8;  DB 3;  Length 2655;
  Best Local Similarity   89.5%;  Pred. No. 1.8e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagcacgttg 25
           |||||||||||  |||| ||
Db      45 GGTGCAGGTCACCACGCTG 27


RESULT  45
US-08-751-767A-7
; Sequence 7, Application US/08751767A
; Patent No. 5994104
;  GENERAL INFORMATION:
;    APPLICANT:  ANDERSON, ROBERT J.
;    APPLICANT:  GRANT, HUGH
;    APPLICANT:  MACDONALD, IAN D.
;    TITLE OF INVENTION:  INTERLUKIN-12 FUSION PROTEIN
;    NUMBER OF SEQUENCES:  80
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  NIXON & VANDERHYE P.C.
;      STREET:  1100 NORTH GLEBE ROAD
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22201
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/751,767A
;      FILING DATE:  08-NOV-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SADOFF, B.J.
;      REGISTRATION NUMBER:  36,663
;      REFERENCE/DOCKET NUMBER:  117-221
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  7038164091
;       TELEFAX:  7038164100
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6139 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  both
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  3222..4841
US-08-751-767A-7


  Query Match             63.2%;  Score 15.8;  DB 2;  Length 6139;
  Best Local Similarity   89.5%;  Pred. No. 1.9e+02;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagc 19
           ||| |||||| ||||||||
Db    3968 TCGGCAGGTGGAGGTCAGC 3986


Search completed: December 26, 2001, 12:29:27
Job time: 7524 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:12:20 ; Search time 4612.2 Seconds
(without alignments)
58.246 Million cell updates/sec

Title: US-09-396-196F-10
Perfect score: 25
Sequence: 1 tcgtcaggtgcaggtcagcacgttg 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	19.2	76.8	895	11	BI155356	BI155356 602903715
c 2	18.6	74.4	416	10	AA146146	AA146146 mq65d12.r
c 3	18.6	74.4	446	10	AA881231	AA881231 vx10d04.r
4	18.6	74.4	618	10	AI945397	AI945397 bs12d09.y
5	18.6	74.4	679	11	BF491517	BF491517 AT28277.5
6	18.6	74.4	688	11	BF504829	BF504829 AT06860.5
c 7	18.6	74.4	765	11	BG480906	BG480906 602530053
c 8	18.6	74.4	815	11	BF313972	BF313972 601901405
9	18.2	72.8	271	10	BB595223	BB595223 BB595223
10	18.2	72.8	318	10	AW710402	AW710402 e3e05ne.r
11	18.2	72.8	370	11	BI068290	BI068290 C019P62U
c 12	18.2	72.8	461	10	AA809928	AA809928 oa93g06.s
c 13	18.2	72.8	480	10	AW710401	AW710401 e3e05ne.f
c 14	18.2	72.8	726	13	AZ656921	AZ656921 1M0532H13
15	18	72.0	1041	13	CNS02BOY	AL190123 Tetraodon
c 16	17.8	71.2	528	13	AZ790140	AZ790140 2M0038E01
c 17	17.8	71.2	730	10	AI478077	AI478077 MPMGp621_
c 18	17.6	70.4	196	11	BI037677	BI037677 MR2-NT013
c 19	17.6	70.4	407	11	BF081725	BF081725 PM0-AN008
20	17.6	70.4	429	10	AW837781	AW837781 CM1-LT004
c 21	17.6	70.4	460	13	AZ441001	AZ441001 1M0232G24
c 22	17.6	70.4	575	11	BF733584	BF733584 PM0-AN008
c 23	17.6	70.4	776	13	AZ933567	AZ933567 BJ__Ba000
c 24	17.6	70.4	792	13	AZ933878	AZ933878 BJ__Ba000
c 25	17.6	70.4	815	11	BI334802	BI334802 602998917
c 26	17.6	70.4	871	11	BF233227	BF233227 602023781
c 27	17.6	70.4	964	13	CNS02KFE	AL201443 Tetraodon
c 28	17.6	70.4	1064	11	BI078629	BI078629 602872730
c 29	17.6	70.4	1077	10	BE620179	BE620179 601483234
c 30	17.6	70.4	1079	13	CNS03FBT	AL241490 Tetraodon
c 31	17.4	69.6	288	10	BB516544	BB516544 BB516544
32	17.4	69.6	398	13	AQ298728	AQ298728 HS_3110_B
c 33	17.4	69.6	767	11	BG703880	BG703880 602686932
c 34	17.4	69.6	1101	13	CNS05186	AL316527 Tetraodon
35	17.2	68.8	251	11	BG561755	BG561755 EtESTed99
c 36	17.2	68.8	303	10	AI755334	AI755334 EtESTea36
c 37	17.2	68.8	379	11	BF331947	BF331947 QV2-BT063
38	17.2	68.8	382	11	BG466683	BG466683 EtESTed36
39	17.2	68.8	392	11	BG884833	BG884833 dac31c11.
40	17.2	68.8	408	11	BG466710	BG466710 EtESTed37
41	17.2	68.8	408	11	BG466711	BG466711 EtESTed37
c 42	17.2	68.8	410	10	AI759462	AI759462 EtESTea22
c 43	17.2	68.8	411	11	BG413233	BG413233 EtESTed18
c 44	17.2	68.8	414	11	BG466993	BG466993 EtESTed31
c 45	17.2	68.8	418	10	AI755467	AI755467 EtESTea37
c 46	17.2	68.8	418	11	BG466484	BG466484 EtESTed23
c 47	17.2	68.8	421	11	BG561583	BG561583 EtESTee01
c 48	17.2	68.8	427	10	AI757999	AI757999 EtESTea35
c 49	17.2	68.8	442	11	BG235712	BG235712 EtESTed11
50	17.2	68.8	443	13	AQ444310	AQ444310 GSSTc0372
c 51	17.2	68.8	466	11	BG516136	BG516136 EtESTed49
c 52	17.2	68.8	478	11	BG516408	BG516408 EtESTed58
c 53	17.2	68.8	494	11	BG561248	BG561248 EtESTed82
c 54	17.2	68.8	495	11	BG466608	BG466608 EtESTed35
c 55	17.2	68.8	497	11	BG466895	BG466895 EtESTed40
56	17.2	68.8	506	11	BG516119	BG516119 EtESTed49
c 57	17.2	68.8	529	11	BG466800	BG466800 EtESTed39
58	17.2	68.8	529	11	BG466826	BG466826 EtESTed39
c 59	17.2	68.8	539	11	BI224545	BI224545 602940122
c 60	17.2	68.8	618	13	AQ659544	AQ659544 Sheared D
c 61	17.2	68.8	626	13	AZ330367	AZ330367 1M0055E10
62	17.2	68.8	677	13	AZ335383	AZ335383 1M0065L14
c 63	17.2	68.8	724	13	AZ876584	AZ876584 2M0191C09
c 64	17.2	68.8	934	11	BI115511	BI115511 602863447
65	17.2	68.8	975	11	BF102149	BF102149 601752642
66	17.2	68.8	1082	13	CNS03NN3	AL252264 Tetraodon
c 67	17	68.0	324	11	BF464374	BF464374 UI-M-CG0p
c 68	17	68.0	339	11	BI042379	BI042379 MR4-NT014
69	17	68.0	371	10	AA751654	AA751654 96AS0395
70	17	68.0	392	10	AW312382	AW312382 4136 MARC
c 71	17	68.0	413	11	BE862745	BE862745 UI-M-BH0-
72	17	68.0	441	10	AI645686	AI645686 vs68d03.x
73	17	68.0	446	10	AA822242	AA822242 vw08b11.r
c 74	17	68.0	495	10	AI789420	AI789420 uk99a06.y
75	17	68.0	504	11	BF409171	BF409171 UI-R-BT1-
c 76	17	68.0	506	10	AW318411	AW318411 um96d05.y
c 77	17	68.0	510	10	AI664441	AI664441 ue62c04.r
78	17	68.0	582	13	AQ344188	AQ344188 RPCI11-12
79	17	68.0	586	10	AW741593	AW741593 ur57d04.x
c 80	17	68.0	597	11	BG922591	BG922591 602821537
81	17	68.0	613	10	AW531885	AW531885 UI-R-BS0-
c 82	17	68.0	628	10	AI037274	AI037274 ud60g08.y
83	17	68.0	631	10	AJ396545	AJ396545 AJ396545
c 84	17	68.0	641	13	AZ426129	AZ426129 1M0206H03
c 85	17	68.0	720	10	BE315518	BE315518 601140784

c 86	17	68.0	747	11	BF385691	BF385691	602047346
87	17	68.0	799	11	BG428201	BG428201	602498855
88	17	68.0	837	10	BE567456	BE567456	601340685
89	17	68.0	861	11	BG417540	BG417540	HVSMEk001
c 90	17	68.0	926	11	BG962572	BG962572	602829953
c 91	17	68.0	947	11	BG283023	BG283023	602406620
92	17	68.0	962	13	CNS042I3	AL271524	Tetraodon
c 93	17	68.0	967	11	BF967116	BF967116	602287648
c 94	17	68.0	993	11	BG824129	BG824129	602729269
c 95	17	68.0	996	13	CNS03F6A	AL241291	Tetraodon
c 96	17	68.0	1008	13	CNS0342I	AL226899	Tetraodon
c 97	17	68.0	1770	11	BG286571	BG286571	602382573
98	16.8	67.2	238	10	AW889758	AW889758	RC1-NT003
c 99	16.8	67.2	289	10	AW368882	AW368882	IL2-HT019
c 100	16.8	67.2	297	10	AW353485	AW353485	34064 MAR

ALIGNMENTS

```
RESULT   1
BI155356/c
LOCUS       BI155356      895 bp    mRNA            EST       05-JUL-2001
DEFINITION  602903715F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5033642 5',
            mRNA sequence.
ACCESSION   BI155356
VERSION     BI155356.1  GI:14615357
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 895)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11093  row: m  column: 03
            High quality sequence stop: 825.
FEATURES             Location/Qualifiers
     source          1. .895
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5033642"
                     /clone_lib="NIH_CGAP_Mam3"
                     /tissue_type="tumor, gross tissue"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert 2 kb. Library constructed by Life
                     Technologies, catalog #12017-018. Investigators providing
                     samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
                     for transgenic model: Xu et al., Nature Genetics 22, 37-43
                     (1999). Note: this is a NCI_CGAP Library."
BASE COUNT      135 a    356 c    278 g    126 t
ORIGIN


  Query Match             76.8%;  Score 19.2;  DB 11;  Length 895;
  Best Local Similarity   87.5%;  Pred. No. 6.5e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacgttg 25
          ||| |||| ||||||||||| ||
Db    817 CGTGAGGTCCAGGTCAGCACGGTG 794
```

```
RESULT   2
AA146146/c
LOCUS       AA146146      416 bp    mRNA            EST       18-FEB-1997
DEFINITION  mq65d12.r1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:583607
            5', mRNA sequence.
ACCESSION   AA146146
VERSION     AA146146.1  GI:1715582
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:358255
            Seq primer: -28M13 rev2 from Amersham.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:583607"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      129 a     83 c    125 g     79 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 10;  Length 416;
  Best Local Similarity   84.0%;  Pred. No. 1e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          || ||  |||||||||||||||| ||
Db    180 TCATCTTGTGCAGGTCAGCACGGTG 156
```

```
RESULT   3
AA881231/c
LOCUS       AA881231      446 bp    mRNA            EST       26-MAR-1998
DEFINITION  vx10d04.r1 Soares_thymus_2NbMT Mus musculus cDNA clone
            IMAGE:1264039 5', mRNA sequence.
ACCESSION   AA881231
```

```
VERSION     AA881231.1  GI:2990541
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:666591
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1264039"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      141 a     87 c    131 g     87 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 10;  Length 446;
  Best Local Similarity   84.0%;  Pred. No. 1.1e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          || ||  |||||||||||||||| ||
Db    165 TCATCTTGTGCAGGTCAGCACGGTG 141


RESULT   4
AI945397
LOCUS       AI945397      618 bp    mRNA            EST       08-JAN-2001
DEFINITION  bs12d09.y1 Drosophila melanogaster adult testis library Drosophila
            melanogaster cDNA clone bs12d09 5', mRNA sequence.
ACCESSION   AI945397
VERSION     AI945397.2  GI:9990738
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 618)
  AUTHORS   Andrews,J., Bouffard,G.G., Cheadle,C., Lu,J., Becker,K.G. and
            Oliver,B.
  TITLE     Gene discovery using computational and microarray analysis of
            transcription in the drosophila melanogaster testis
  JOURNAL   Genome Res. 10 (12), 2030-2043 (2000)
  MEDLINE   20568492
COMMENT     On Aug 17, 1999 this sequence version replaced gi:5735795.
            Contact: Brian Oliver
            Laboratory of Cellular and Developmental Biology
            NIDDK, National Institutes of Health
            6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
            Fax: (301) 496 5239
            Email: oliver@helix.nih.gov,
            http://www.niddk.nih.gov/intram/people/boliver.htm
            Tissue isolation and library construction performed at the National
            Institute of Diabetes and Digestive and Kidney Diseases, NIH (see
            http://www.niddk.nih.gov/intram/people/boliver.htm). DNA sequencing
            and analyses performed by National Institutes of Health Intramural
            Sequencing Center (NISC; see http://www.nisc.nih.gov).
            Plate: 12  row: d  column: 09
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .618
                     /organism="Drosophila melanogaster"
                     /strain="y[*] w[67c1]/Y"
                     /db_xref="taxon:7227"
                     /clone="bs12d09"
                     /clone_lib="Drosophila melanogaster adult testis library"
                     /sex="male"
                     /dev_stage="1-5 day adult"
                     /lab_host="SOLR (Stratagene)"
                     /note="Organ: testis; Vector: pBlueScript SK (Stratagene);
                     Site_1: EcoR I; Site_2: Xho I; Testes dissected from 1-5
                     day adult y[*] w[67c1]/Y males raised at 25oC.  RNA
                     isolated using Trizol (Life Technologies) and a single
                     round of Poly(A)+ selection using Oligotex (Qiagen).  cDNA
                     library constructed using Stratagene ZAP-cDNA syntesis
                     kit.  Oligo dT-primed, size fractionated ~1-6 kb, and
                     directionally cloned at EcoRI and XhoI in Uni-ZAP XR.
                     Following a single round of amplification pBlueScript SK
                     phagemids were mass excised.  A distribution channel for
                     clones is being sought, but not currently available.
                     Requests for clones cannot be honored."
BASE COUNT      138 a    180 c    166 g    134 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 10;  Length 618;
  Best Local Similarity   84.0%;  Pred. No. 1.1e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          ||| |||||||||||| ||| | ||
Db    213 TCGACAGGTGCAGGTCTGCAAGGTG 237


RESULT   5
BF491517
LOCUS       BF491517      679 bp    mRNA            EST       19-APR-2001
DEFINITION  AT28277.5prime AT Drosophila melanogaster adult testes pOTB7
            Drosophila melanogaster cDNA clone AT28277 5 similar to CG3515:
            FBan0003515 located on: 2L 22F4-22F4;: 04/09/2001, mRNA sequence.
ACCESSION   BF491517
VERSION     BF491517.2  GI:13695044
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 679)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E., Berman
            ,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V., Farfan
```

```
           ,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P.,
           Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J.,
           Paragas,V., Park,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E.,
           Celniker,S. and Rubin,G.M.
  TITLE    BDGP/HHMI AT Drosophila EST Project
  JOURNAL  Unpublished (2000)
COMMENT    On Dec 6, 2000 this sequence version replaced gi:11574818.
           Contact: Stapleton, M.
           BDGP
           Lawrence Berkeley National Lab
           One Cyclotron Rd, Berkeley, CA 94720, USA
           Fax: 510 486 6798
           Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
           hit genomic AE003582: arm:2L [2452680,2747991]
           estimated-cyto:22F3-23B3: 04/09/2001
           Plate: AT.282  row: G  column: 5
           High quality sequence stop: 650.
FEATURES             Location/Qualifiers
     source          1. .679
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="AT28277"
                     /clone_lib="AT Drosophila melanogaster adult testes pOTB7"
                     /sex="male"
                     /dev_stage="0-3 day old Ore-R males"
                     /lab_host="Plates AT.10-AT.120: DH5-alpha. Plates
                     AT.121-AT.319: DH5-alpha TonA"
                     /note="Organ: ADULT testes; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; The mRNA for the testis library was made
                     from testes and seminal vessicles hand dissected from 0-3
                     day old Ore-R males. RNA kindly provided by the lab of
                     Margaret Fuller. Sized fractionated cDNAs were directly
                     ligated into pOTB7. Plasmid cDNA library."
BASE COUNT      161 a    190 c    178 g    150 t
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 11;  Length 679;
  Best Local Similarity   84.0%;  Pred. No. 1.1e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           ||| |||||||||||| ||| | ||
Db     246 TCGACAGGTGCAGGTCTGCAAGGTG 270


RESULT   6
BF504829
LOCUS       BF504829      688 bp    mRNA            EST       19-APR-2001
DEFINITION  AT06860.5prime AT Drosophila melanogaster adult testes pOTB7
            Drosophila melanogaster cDNA clone AT06860 5 similar to CG3515:
            FBan0003515 located on: 2L 22F4-22F4;: 04/07/2001, mRNA sequence.
ACCESSION   BF504829
VERSION     BF504829.2  GI:13687447
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 688)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E., Berman
            ,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V., Farfan
            ,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P.,
            Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J.,
            Paragas,V., Park,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E.,
            Celniker,S. and Rubin,G.M.
  TITLE     BDGP/HHMI AT Drosophila EST Project
  JOURNAL   Unpublished (2000)
COMMENT     On Dec 6, 2000 this sequence version replaced gi:11588130.
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003582: arm:2L [2452680,2747991]
            estimated-cyto:22F3-23B3: 04/07/2001
            Plate: AT.68  row: E  column: 12
            High quality sequence stop: 559.
FEATURES             Location/Qualifiers
     source          1. .688
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="AT06860"
                     /clone_lib="AT Drosophila melanogaster adult testes pOTB7"
                     /sex="male"
                     /dev_stage="0-3 day old Ore-R males"
                     /lab_host="Plates AT.10-AT.120: DH5-alpha. Plates
                     AT.121-AT.319: DH5-alpha TonA"
                     /note="Organ: ADULT testes; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; The mRNA for the testis library was made
                     from testes and seminal vessicles hand dissected from 0-3
                     day old Ore-R males. RNA kindly provided by the lab of
                     Margaret Fuller. Sized fractionated cDNAs were directly
                     ligated into pOTB7. Plasmid cDNA library."
BASE COUNT      157 a    198 c    179 g    153 t      1 others
ORIGIN


  Query Match             74.4%;  Score 18.6;  DB 11;  Length 688;
  Best Local Similarity   84.0%;  Pred. No. 1.1e+03;
  Matches   21;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           ||| |||||||||||| ||| | ||
Db     225 TCGACAGGTGCAGGTCTGCAAGGTG 249


RESULT   7
BG480906/c
LOCUS       BG480906      765 bp    mRNA            EST       21-MAR-2001
DEFINITION  602530053F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4653420 5',
            mRNA sequence.
ACCESSION   BG480906
VERSION     BG480906.1  GI:13413185
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 765)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1439  row: j  column: 13
            High quality sequence stop: 621.
FEATURES             Location/Qualifiers
     source          1. .765
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4653420"
                     /clone_lib="NIH_MGC_21"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
```

Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 182 a 191 c 224 g 168 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 11; Length 765;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcacgttg 25
||||||||| |||||||| |||
Db 163 TCGTCAGGTGAGAGTCAGCACCTTG 139

RESULT 8
BF313972/c
LOCUS BF313972 815 bp mRNA EST 21-NOV-2000
DEFINITION 601901405F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4130760 5', mRNA sequence.
ACCESSION BF313972
VERSION BF313972.1 GI:11262070
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 815)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM1027 row: i column: 01
High quality sequence stop: 698.
FEATURES Location/Qualifiers
source 1. .815
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4130760"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 231 a 178 c 251 g 155 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 11; Length 815;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcacgttg 25
||||||||| |||||||| |||
Db 708 TCGTCAGGTGAGAGTCAGCACCTTG 684

RESULT 9
BB595223
LOCUS BB595223 271 bp mRNA EST 30-NOV-2000
DEFINITION BB595223 RIKEN full-length enriched, adult male corpus striatum Mus musculus cDNA clone C030005E12 5', mRNA sequence.
ACCESSION BB595223
VERSION BB595223.1 GI:11491825
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 271)
AUTHORS Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodoyama,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka,T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .271
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C030005E12"
/clone_lib="RIKEN full-length enriched, adult male corpus striatum"
/sex="male"
/tissue_type="corpus striatum"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second

strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
BASE COUNT 62 a 63 c 95 g 51 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 271;
Best Local Similarity 87.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 gtcaggtgcaggtcagcacgttg 25
| |||||| ||||||||| ||||
Db 244 GCCAGGTGGAGGTCAGCAGGTTG 266

RESULT 10
AW710402
LOCUS AW710402 318 bp mRNA EST 25-APR-2000
DEFINITION e3e05ne.r1 Neurospora crassa evening cDNA library Neurospora crassa cDNA clone e3e05ne 3', mRNA sequence.
ACCESSION AW710402
VERSION AW710402.1 GI:7599490
KEYWORDS EST.
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariales; Sordariaceae; Neurospora.
REFERENCE 1 (bases 1 to 318)
AUTHORS Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.
TITLE Two Neurospora crassa EST Databases
JOURNAL Unpublished (1998)
COMMENT Other_ESTs: e3e05ne.f1
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
Seq primer: Universal Reverse Primer
High quality sequence stop: 80.
FEATURES Location/Qualifiers
source 1. .318
/organism="Neurospora crassa"
/strain="Strain 30-7 (bd; A)"
/db_xref="taxon:5141"
/clone="e3e05ne"
/clone_lib="Neurospora crassa evening cDNA library"
/tissue_type="tissue harvested following 22hr growth in dark"
/note="Vector: pBlueScript SK-; Site_1: XbaI; Site_2: EcoRI; See: Bell-Perdersen,D., et al. PNAS 93:13096,1996. 5' end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript"
BASE COUNT 62 a 81 c 67 g 108 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 318;
Best Local Similarity 87.0%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cgtcaggtgcaggtcagcacgtt 24
|||||||||| | | ||||||||
Db 99 CGTCAGGTGCTGCTTAGCACGTT 121

RESULT 11
BI068290
LOCUS BI068290 370 bp mRNA EST 19-JUN-2001
DEFINITION C019P62U Populus strain T89 leaves Populus tremula x Populus tremuloides cDNA, mRNA sequence.
ACCESSION BI068290
VERSION BI068290.1 GI:14488910
KEYWORDS EST.
SOURCE Populus tremula x Populus tremuloides.
ORGANISM Populus tremula x Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE 1 (bases 1 to 370)
AUTHORS Jansson,S., Bhalerao,R., Erlandsson,R., Bjorkbacka,H., Karlsson,J., Sterky,F., Gustafsson,P. and Lundeberg,J.
TITLE Gene expression in Populus leaves
JOURNAL Unpublished (2001)
COMMENT Contact: Erlandsson R
Department of Biotechnology
Royal Institute of Technology
Teknikringen 30, Stockholm S-10044, Sweden
Tel: 46 8 790 8287
Fax: 46 8 245452
Email: rikerl@biochem.kth.se.
FEATURES Location/Qualifiers
source 1. .370
/organism="Populus tremula x Populus tremuloides"
/strain="T89"
/db_xref="taxon:47664"
/clone_lib="Populus strain T89 leaves"
/tissue_type="leaf"
BASE COUNT 85 a 83 c 94 g 101 t 7 others
ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 370;
Best Local Similarity 83.3%; Pred. No. 1.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcacgtt 24
|||||| |||||| || ||||||
Db 218 TCGTCAAGTGCAGTNCATCACGTT 241

RESULT 12
AA809928/c
LOCUS AA809928 461 bp mRNA EST 19-FEB-1998
DEFINITION oa93g06.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1319866 3' similar to gb:X74301_cds1 MHC CLASS II TRANSACTIVATOR CIITA (HUMAN);; mRNA sequence.
ACCESSION AA809928
VERSION AA809928.1 GI:2879334
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 461)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 904 Std Error: 0.00

```
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 221.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1319866"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT       94 a    140 c    133 g     93 t      1 others
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 461;
  Best Local Similarity   87.0%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgt 23
           ||||| ||||| |||| |||||||
Db     276 TCGTCCGGTGCCGGTCCGCACGT 254


RESULT  13
AW710401/c
LOCUS       AW710401      480 bp    mRNA            EST       25-APR-2000
DEFINITION  e3e05ne.f1 Neurospora crassa evening cDNA library Neurospora crassa
            cDNA clone e3e05ne 5', mRNA sequence.
ACCESSION   AW710401
VERSION     AW710401.1  GI:7599489
KEYWORDS    EST.
SOURCE      Neurospora crassa.
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariales; Sordariaceae; Neurospora.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.
  TITLE     Two Neurospora crassa EST Databases
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: e3e05ne.r1
            Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
            Department of Chemistry and Biochemistry
            Advanced Center for Genome Technology, University of Oklahoma
            620 Parrington Oval, Norman, OK 73019, USA
            Tel: 405 325 4912
            Fax: 405 325 7762
            Email: broe@ou.edu
            We anticipate the future release of the cDNA clones to the Fungal
            Genetics Stock Center
            Seq primer: Universal Forward Primer
            High quality sequence stop: 401.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Neurospora crassa"
                     /strain="Strain 30-7 (bd; A)"
                     /db_xref="taxon:5141"
                     /clone="e3e05ne"
                     /clone_lib="Neurospora crassa evening cDNA library"
                     /tissue_type="tissue harvested following 22hr growth in
                     dark"
                     /note="Vector: pBlueScript SK-; Site_1: XbaI; Site_2:
                     EcoRI; See: Bell-Perdersen,D., et al. PNAS 93:13096,1996.
                     5' end of cDNA cloned into XbaI site of pBluescript; 3'
                     end of cDNA cloned into EcoRI site of pBluescript"
BASE COUNT      146 a    121 c    127 g     86 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 10;  Length 480;
  Best Local Similarity   87.0%;  Pred. No. 1.5e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgtt 24
           ||||||||||| | | ||||||||
Db     431 CGTCAGGTGCTGCTTAGCACGTT 409


RESULT  14
AZ656921/c
LOCUS       AZ656921      726 bp    DNA             GSS       14-DEC-2000
DEFINITION  1M0532H13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0532H13 R, DNA sequence.
ACCESSION   AZ656921
VERSION     AZ656921.1  GI:11794067
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 726)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0532  row: H  column: 13
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 726.
FEATURES             Location/Qualifiers
     source          1. .726
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0532H13"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
```

```
                    with adaptors complementary to the insert adaptors and
                    purified. The sheared, adaptored mouse DNA was annealed to
                    adaptored vector DNA, and transformed into
                    chemically-competent E. coli XL10-Gold (Stratagene) cells
                    and selected for ampicillin resistance."
BASE COUNT      207 a    187 c    143 g    189 t
ORIGIN


  Query Match             72.8%;  Score 18.2;  DB 13;  Length 726;
  Best Local Similarity   87.0%;  Pred. No. 1.6e+03;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcagcacgttg 25
           |||||||| ||||||| || |||
Db     674 GTCAGGTGAAGGTCAGAACATTG 652


RESULT  15
CNS02BOY
LOCUS       CNS02BOY     1041 bp    DNA             GSS       12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            253N02 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL190123
VERSION     AL190123.1  GI:7828227
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1041)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="253N02"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG253DG01LP1~end : T7"
BASE COUNT      253 a    301 c    285 g    180 t     22 others
ORIGIN


  Query Match             72.0%;  Score 18;  DB 13;  Length 1041;
  Best Local Similarity   90.0%;  Pred. No. 1.9e+03;
  Matches   18;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcac 21
           | ||||||||||||:|||||
Db      92 CYTCAGGTGCAGGTYAGCAC 111


RESULT  16
AZ790140/c
LOCUS       AZ790140      528 bp    DNA             GSS       16-FEB-2001
DEFINITION  2M0038E01R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0038E01 R, DNA sequence.
ACCESSION   AZ790140
VERSION     AZ790140.1  GI:12931689
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0038  row: E  column: 01
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 528.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0038E01"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      154 a    111 c    114 g    149 t
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 13;  Length 528;
  Best Local Similarity   90.5%;  Pred. No. 2.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtcagcacgtt 24
           |||||||||||| |||||| |
```

```
Db    366 TCAGGTGCAGGTGAGCACGAT 346


RESULT 17
AI478077/c
LOCUS       AI478077      730 bp    mRNA            EST       09-AUG-1999
DEFINITION  MPMGp621_042D15 MPMGp621 Strongylocentrotus purpuratus cDNA clone
            042D15 similar to beta-catenin, mRNA sequence.
ACCESSION   AI478077
VERSION     AI478077.1  GI:5713397
KEYWORDS    EST.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 730)
  AUTHORS   Poustka,A.J., Herwig,R., Krause,A., Hennig,S., Meier-Ewert,S. and
            Lehrach,H.
  TITLE     Towards the gene catalogue of sea urchin development: The
            construction and analysis of an unfertilized egg cDNA library
            highly normalized by oligonucleotide fingerprinting
  JOURNAL   Genomics 59 (2), 122-133 (1999)
  MEDLINE   99339976
COMMENT     Contact: Poustka A.J.
            Department Lehrach
            Max-Planck-Institute for Molecular Genetics
            Ihnestrasse 73, 14195 Berlin, Germany
            Tel: +49-30-84131623
            Fax: +49-30-84131128
            Email: poustka@mpimg-berlin-dahlem.mpg.de
            Insert Length: 1500   Std Error: 0.00
            High quality sequence start: 53
            High quality sequence stop: 541.
FEATURES             Location/Qualifiers
     source          1. .730
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="042D15"
                     /clone_lib="MPMGp621"
                     /tissue_type="unfertilized egg"
                     /note="Vector: pSport1; oligo(dT) primed, directionally
                     cloned library"
BASE COUNT      167 a    208 c    165 g    182 t      8 others
ORIGIN


  Query Match             71.2%;  Score 17.8;  DB 10;  Length 730;
  Best Local Similarity   90.5%;  Pred. No. 2.2e+03;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagcacgttg 25
           |||| |||||||| |||||||
Db     370 CAGGCGCAGGTCACCACGTTG 350


RESULT 18
BI037677/c
LOCUS       BI037677      196 bp    mRNA            EST       14-JUN-2001
DEFINITION  MR2-NT0138-110101-010-b03_1 NT0138 Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   BI037677
VERSION     BI037677.1  GI:14444303
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR2&t2=MR2-NT0138-
            110101-010-b03_1&t3=2001-01-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 56.
FEATURES             Location/Qualifiers
     source          1. .196
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0138"
                     /dev_stage="Adult"
                     /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       39 a     57 c     55 g     44 t      1 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 196;
  Best Local Similarity   83.3%;  Pred. No. 2.4e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           ||  ||||| ||||||||||| ||
Db     182 CGATAGGTGGAGGTCAGCACGGTG 159


RESULT 19
BF081725/c
LOCUS       BF081725      407 bp    mRNA            EST       18-OCT-2000
DEFINITION  PM0-AN0087-130900-003-a11 AN0087 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF081725
VERSION     BF081725.1  GI:10875555
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
```

```
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=PM0-AN0087-130
            900-003-a11&t3=2000-09-13&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 12
            High quality sequence stop: 407.
FEATURES              Location/Qualifiers
     source           1. .407
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="AN0087"
                      /dev_stage="Adult"
                      /note="Organ: amnion_normal; Vector: puc18; Site_1: SmaI;
                      Site_2: SmaI; A mini-library was made by cloning products
                      derived from  ORESTES PCR (U.S. Letters Patent application
                      No. 196,716 -  Ludwig Institute for Cancer Research)
                      profiles into  the pUC 18 vector. Reverse transcription of
                      tissue mRNA and cDNA amplification were performed under
                      low stringency conditions."
BASE COUNT      109 a     96 c    113 g     88 t      1 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 407;
  Best Local Similarity   83.3%;  Pred. No. 2.6e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           | ||||||||||| || |||| |
Db     146 CTTCAGGTGCAGGTAAGTACGTGG 123


RESULT  20
AW837781
LOCUS       AW837781      429 bp    mRNA            EST       18-MAY-2000
DEFINITION  CM1-LT0042-100300-140-b12 LT0042 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW837781
VERSION     AW837781.1  GI:7931755
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM1-LT0042-100
            300-140-b12&t3=2000-03-10&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 9
            High quality sequence stop: 429.
FEATURES              Location/Qualifiers
     source           1. .429
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="LT0042"
                      /dev_stage="Adult"
                      /note="Organ: leiomios; Vector: puc18; Site_1: SmaI;
                      Site_2: SmaI; A mini-library was made by cloning products
                      derived from  ORESTES PCR (U.S. Letters Patent application
                      No. 196,716 -  Ludwig Institute for Cancer Research)
                      profiles into  the pUC 18 vector. Reverse transcription of
                      tissue mRNA and cDNA amplification were performed under
                      low stringency conditions."
BASE COUNT       73 a    132 c     92 g    132 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 429;
  Best Local Similarity   83.3%;  Pred. No. 2.6e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgtt 24
           || ||||||||||||| | || ||
Db      58 TCTTCAGGTGCAGGTCTGGACCTT 81


RESULT  21
AZ441001/c
LOCUS       AZ441001      460 bp    DNA             GSS       03-OCT-2000
DEFINITION  1M0232G24F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0232G24 F, DNA sequence.
ACCESSION   AZ441001
VERSION     AZ441001.1  GI:10565014
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0232  row: G  column: 24
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 460.
FEATURES              Location/Qualifiers
     source           1. .460
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0232G24"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
```

0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

```
BASE COUNT      110 a    127 c    120 g    103 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 460;
  Best Local Similarity   83.3%;  Pred. No. 2.6e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           | ||||| |||  ||||||||||||
Db     173 CTTCAGGCGCATCTCAGCACGTTG 150


RESULT  22
BF733584/c
LOCUS       BF733584      575 bp    mRNA            EST       09-JAN-2001
DEFINITION  PM0-AN0087-091000-005-e02 AN0087 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF733584
VERSION     BF733584.1  GI:12058820
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 575)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM0&t2=PM0-AN0087-
            091000-005-e02&t3=2000-10-09&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 574.
FEATURES             Location/Qualifiers
     source          1. .575
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="AN0087"
                     /dev_stage="Adult"
                     /note="Organ: amnion_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      159 a    146 c    157 g    112 t      1 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 575;
  Best Local Similarity   83.3%;  Pred. No. 2.6e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           | ||||||||||||| || |||| |
Db     135 CTTCAGGTGCAGGTAAGTACGTGG 112


RESULT  23
AZ933567/c
LOCUS       AZ933567      776 bp    DNA             GSS       24-APR-2001
DEFINITION  BJ__Ba0001A17f B. japonicum BAC library Bradyrhizobium japonicum
            genomic, DNA sequence.
ACCESSION   AZ933567
VERSION     AZ933567.1  GI:13775627
KEYWORDS    GSS.
SOURCE      Bradyrhizobium japonicum.
  ORGANISM  Bradyrhizobium japonicum
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Bradyrhizobium group; Bradyrhizobium.
REFERENCE   1  (bases 1 to 776)
  AUTHORS   Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Goicoechea
            ,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
  TITLE     A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
            genome
  JOURNAL   Genome Res. 11 (8), 1434-1440 (2001)
  MEDLINE   21376150
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Class: BAC ends
            High quality sequence stop: 723.
FEATURES             Location/Qualifiers
     source          1. .776
                     /organism="Bradyrhizobium japonicum"
                     /strain="USDA110"
                     /db_xref="taxon:375"
                     /clone_lib="B. japonicum BAC library"
                     /lab_host="E. coli"
                     /note="Vector: pIndigo536; Site_1: HindIII"
BASE COUNT      139 a    245 c    246 g    141 t      5 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 776;
  Best Local Similarity   83.3%;  Pred. No. 2.7e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           || | || |||||||||||||| ||
Db     146 CGGCCGGCGCAGGTCAGCACGATG 123


RESULT  24
AZ933878/c
LOCUS       AZ933878      792 bp    DNA             GSS       24-APR-2001
DEFINITION  BJ__Ba0001O02f B. japonicum BAC library Bradyrhizobium japonicum
            genomic, DNA sequence.
ACCESSION   AZ933878
VERSION     AZ933878.1  GI:13775938
```

```
KEYWORDS    GSS.
SOURCE      Bradyrhizobium japonicum.
  ORGANISM  Bradyrhizobium japonicum
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Bradyrhizobium group; Bradyrhizobium.
REFERENCE   1  (bases 1 to 792)
  AUTHORS   Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Goicoechea
            ,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
  TITLE     A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
            genome
  JOURNAL   Genome Res. 11 (8), 1434-1440 (2001)
  MEDLINE   21376150
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Class: BAC ends
            High quality sequence stop: 748.
FEATURES             Location/Qualifiers
     source          1. .792
                     /organism="Bradyrhizobium japonicum"
                     /strain="USDA110"
                     /db_xref="taxon:375"
                     /clone_lib="B. japonicum BAC library"
                     /lab_host="E. coli"
                     /note="Vector: pIndigo536; Site_1: HindIII"
BASE COUNT      139 a    251 c    244 g    156 t      2 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 792;
  Best Local Similarity   83.3%;  Pred. No. 2.7e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacgttg 25
          || | || |||||||||||||| ||
Db    146 CGGCCGGCGCAGGTCAGCACGATG 123


RESULT  25
BI334802/c
LOCUS       BI334802      815 bp    mRNA            EST       30-JUL-2001
DEFINITION  602998917F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5140901 5',
            mRNA sequence.
ACCESSION   BI334802
VERSION     BI334802.1  GI:15019459
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 815)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11347  row: b  column: 06
            High quality sequence stop: 597.
FEATURES             Location/Qualifiers
     source          1. .815
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5140901"
                     /clone_lib="NIH_MGC_12"
                     /tissue_type="cervical carcinoma cell line"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.4 kb. Library prepared by Life
                     Technologies."
BASE COUNT      170 a    304 c    191 g    149 t      1 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 815;
  Best Local Similarity   83.3%;  Pred. No. 2.7e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacgttg 25
          || ||| |||||||||||| || ||
Db    512 CGCCAGTTGCAGGTCAGCCCGCTG 489


RESULT  26
BF233227/c
LOCUS       BF233227      871 bp    mRNA            EST       14-NOV-2000
DEFINITION  602023781F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4159005 5',
            mRNA sequence.
ACCESSION   BF233227
VERSION     BF233227.1  GI:11143414
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 871)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9436  row: a  column: 22
            High quality sequence stop: 626.
FEATURES             Location/Qualifiers
     source          1. .871
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4159005"
                     /clone_lib="NCI_CGAP_Li9"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      222 a    282 c    226 g    141 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 871;
  Best Local Similarity   83.3%;  Pred. No. 2.7e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgtt 24
          ||| |||||| |||||||  ||||
Db    535 TCGGCAGGTGTAGGTCAGGTCGTT 512
```

```
RESULT  27
CNS02KFE/c
LOCUS       CNS02KFE      964 bp    DNA              GSS       14-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            145K06 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL201443
VERSION     AL201443.1  GI:7859788
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 964)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 964)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 964)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .964
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="145K06"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG145BF03SP1~end :
                     PUC-Ori"
BASE COUNT      279 a    208 c    203 g    259 t     15 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 964;
  Best Local Similarity   83.3%;  Pred. No. 2.8e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacgttg 25
           |||||||||   |||||||||| |
Db     665 CGTCAGGTGGCTGTCAGCACGTGG 642


RESULT  28
BI078629/c
LOCUS       BI078629     1064 bp    mRNA             EST       20-JUN-2001
DEFINITION  602872730F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5004798 5',
            mRNA sequence.
ACCESSION   BI078629
VERSION     BI078629.1  GI:14496959
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1064)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11043  row: k  column: 07
            High quality sequence start: 10
            High quality sequence stop: 418.
FEATURES             Location/Qualifiers
     source          1. .1064
                     /organism="Mus musculus"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5004798"
                     /clone_lib="NCI_CGAP_Mam2"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      274 a    332 c    283 g    175 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 11;  Length 1064;
  Best Local Similarity   83.3%;  Pred. No. 2.8e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgtt 24
           || |||||||||||||| | | |
Db     210 TCCTCAGGTGCAGGTCAGTATGGT 187


RESULT  29
BE620179/c
LOCUS       BE620179     1077 bp    mRNA             EST       20-OCT-2000
DEFINITION  601483234F1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3885657 5',
            mRNA sequence.
ACCESSION   BE620179
VERSION     BE620179.1  GI:9891117
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1077)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9661  row: d  column: 10
            High quality sequence stop: 643.
FEATURES             Location/Qualifiers
     source          1. .1077
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3885657"
                     /clone_lib="NIH_MGC_69"
```

```
                     /tissue_type="large cell carcinoma, undifferentiated"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.1 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      324 a    278 c    322 g    153 t
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 10;  Length 1077;
  Best Local Similarity   83.3%;  Pred. No. 2.8e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcagcacgttg 25
          | ||||||||||||  || |||| |
Db    228 CTTCAGGTGCAGGTAAGTACGTGG 205


RESULT  30
CNS03FBT/c
LOCUS       CNS03FBT     1079 bp    DNA             GSS       17-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            021L24 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL241490
VERSION     AL241490.1  GI:7962259
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1079)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1079)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1079)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1079
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="021L24"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG021DF12SP1-end :
                     PUC-Ori"
BASE COUNT      263 a    247 c    279 g    288 t      2 others
ORIGIN


  Query Match             70.4%;  Score 17.6;  DB 13;  Length 1079;
  Best Local Similarity   83.3%;  Pred. No. 2.8e+03;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgtt 24
          |||| ||||| | |||||||| ||
Db   1036 TCGTGAGGTGAAAGTCAGCACTTT 1013


RESULT  31
BB516544/c
LOCUS       BB516544      288 bp    mRNA            EST       28-JUL-2000
DEFINITION  BB516544 RIKEN full-length enriched, 16 days neonate heart Mus
            musculus cDNA clone D830016J11 3' similar to M76601 Mouse alpha
            cardiac myosin heavy chain mRNA, mRNA sequence.
ACCESSION   BB516544
VERSION     BB516544.1  GI:9568002
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 288)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .288
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D830016J11"
                     /clone_lib="RIKEN full-length enriched, 16 days neonate
                     heart"
                     /tissue_type="heart"
                     /dev_stage="16 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
```

GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 83 a 104 c 57 g 44 t
ORIGIN

Query Match 69.6%; Score 17.4; DB 10; Length 288;
Best Local Similarity 94.7%; Pred. No. 3e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 cgtcaggtgcaggtcagca 20
||||||||||||||| ||
Db 188 CGTCAGGTGCAGGTCATCA 170

RESULT 32
AQ298728
LOCUS AQ298728 398 bp DNA GSS 15-DEC-1998
DEFINITION HS_3110_B1_C12_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3110 Col=23 Row=F, DNA sequence.
ACCESSION AQ298728
VERSION AQ298728.1 GI:4015907
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 398)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3110 row: F column: 23
Class: BAC ends
High quality sequence stop: 398.
FEATURES Location/Qualifiers
source 1. .398
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3110 Col=23 Row=F"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 87 a 93 c 102 g 116 t
ORIGIN

Query Match 69.6%; Score 17.4; DB 13; Length 398;
Best Local Similarity 94.7%; Pred. No. 3.1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 7 ggtgcaggtcagcacgttg 25
||||||||||||||| ||
Db 330 GGTGCAGGTCAGCACGATG 348

RESULT 33
BG703880/c
LOCUS BG703880 767 bp mRNA EST 07-MAY-2001
DEFINITION 602686932F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4819446 5', mRNA sequence.
ACCESSION BG703880
VERSION BG703880.1 GI:13976654
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 767)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10722 row: p column: 07
High quality sequence stop: 536.
FEATURES Location/Qualifiers
source 1. .767
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4819446"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 139 a 215 c 202 g 211 t
ORIGIN

Query Match 69.6%; Score 17.4; DB 11; Length 767;
Best Local Similarity 94.7%; Pred. No. 3.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 cgtcaggtgcaggtcagca 20
||||||||||||||| |
Db 649 CGTCAGGTGCAGGTCAGTA 631

RESULT 34
CNS05186/c
LOCUS CNS05186 1101 bp DNA GSS 26-JUL-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone 025P14 of library A from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL316527
VERSION AL316527.1 GI:9549411
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

```
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
REFERENCE   2  (bases 1 to 1101)
  AUTHORS   Crollius,H.R., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
            Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
            Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
REFERENCE   3  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="025P14"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA025DH07A1-end : T3"
BASE COUNT      255 a    267 c    271 g    274 t     34 others
ORIGIN


  Query Match             69.6%;  Score 17.4;  DB 13;  Length 1101;
  Best Local Similarity   94.7%;  Pred. No. 3.3e+03;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtcagcacg 22
          |||||||||||||| ||||
Db    759 TCAGGTGCAGGTCACCACG 741


RESULT  35
BG561755
LOCUS       BG561755      251 bp    mRNA            EST       10-APR-2001
DEFINITION  EtESTed99h03.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5', mRNA sequence.
ACCESSION   BG561755
VERSION     BG561755.1  GI:13590753
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 251)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .251
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                     /dev_stage="Sporozoite stage"
                     /lab_host="SOLR"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E.tenella grown in chickens. cDNA was
                     synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to the cDNA and products were
                     size-selected on sephacryl S500. The cDNA were ligated to
                     EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
                     converted to phagemids by mass excision using ExAssist
                     helper phage and E.coli SOLR cell (Stratagene). Clones
                     were selected by negative hybridization against a pool of
                     overrepresented ESTs (N>=10,from 1682 previous reads).
                     Insert sizes range from 1.2-2.9Kb. The library may contain
                     a small percentage of host or bacterial contaminents."
BASE COUNT       59 a     52 c     67 g     73 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 11;  Length 251;
  Best Local Similarity   86.4%;  Pred. No. 3.6e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtcagcacgttg 25
          ||| |||||||||| ||| |||
Db    141 TCATGTGCAGGTCATCACATTG 162


RESULT  36
AI755334/c
LOCUS       AI755334      303 bp    mRNA            EST       18-JAN-2000
DEFINITION  EtESTea36c11.y1 Eimeria S5-2 Sporozoite stage Eimeria tenella cDNA
            5', mRNA sequence.
ACCESSION   AI755334
VERSION     AI755334.1  GI:5149133
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 303)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 280.
FEATURES             Location/Qualifiers
     source          1. .303
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
```

/clone_lib="Eimeria S5-2 Sporozoite stage"
/dev_stage="Sporozoite"
/lab_host="SOLR E. coli"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI ; Sporozoites were obtained from in vitro sporulated and excysted oocysts of E. tenella grown in chickens. cDNA was synthesized from poly mRNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on Sephacryl S500. cDNAs were digested with EcoRI/XhoI and cloned into lambda Zap II (Stratagene). Clones were converted to phagemids by mass excision using ExAssist helper phage and SOLR cells (Stratagene). Insert sizes range from 1.2-2.9 kb."
BASE COUNT 87 a 80 c 58 g 77 t 1 others
ORIGIN

Query Match 68.8%; Score 17.2; DB 10; Length 303;
Best Local Similarity 86.4%; Pred. No. 3.6e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 tcaggtgcaggtcagcacgttg 25
||| |||||||||| ||| |||
Db 129 TCATGTGCAGGTCATCACATTG 108

RESULT 37
BF331947/c
LOCUS BF331947 379 bp mRNA EST 22-NOV-2000
DEFINITION QV2-BT0634-280800-331-e08 BT0634 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF331947
VERSION BF331947.1 GI:11302695
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 379)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV2&t2=QV2-BT0634-280800-331-e08&t3=2000-08-28&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 217.
FEATURES Location/Qualifiers
source 1. .379
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0634"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 ,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 127 a 59 c 88 g 105 t
ORIGIN

Query Match 68.8%; Score 17.2; DB 11; Length 379;
Best Local Similarity 86.4%; Pred. No. 3.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcacg 22
||| |||||||||| || |||||
Db 241 TCGACAGGTGCAGCTCTGCACG 220

RESULT 38
BG466683
LOCUS BG466683 382 bp mRNA EST 20-MAR-2001
DEFINITION EtESTed36g01.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria tenella cDNA 5', mRNA sequence.
ACCESSION BG466683
VERSION BG466683.1 GI:13395658
KEYWORDS EST.
SOURCE Eimeria tenella.
ORGANISM Eimeria tenella
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae; Eimeria.
REFERENCE 1 (bases 1 to 382)
AUTHORS Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T., Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
TITLE WashU-Merck Eimeria tenella project
JOURNAL Unpublished (1999)
COMMENT Contact: David Sibley, Ph.D.
WashU-Merck Eimeria tenella project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Contact David Sibley (toxoest@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco.
FEATURES Location/Qualifiers
source 1. .382
/organism="Eimeria tenella"
/strain="LS18"
/db_xref="taxon:5802"
/clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
/dev_stage="Sporozoite stage"
/lab_host="SOLR"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI ; Sporozoites were obtained from in vitro sporulated and excysted oocysts of E.tenella grown in chickens. cDNA was synthesized from poly mRNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on sephacryl S500. The cDNA were ligated to EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were converted to phagemids by mass excision using ExAssist helper phage and E.coli SOLR cell (Stratagene). Clones were selected by negative hybridization against a pool of overrepresented ESTs (N>=10,from 1682 previous reads). Insert sizes range from 1.2-2.9Kb. The library may contain a small percentage of host or bacterial contaminents."
BASE COUNT 92 a 74 c 108 g 108 t
ORIGIN

Query Match 68.8%; Score 17.2; DB 11; Length 382;

```
  Best Local Similarity   86.4%;  Pred. No. 3.7e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtcagcacgttg 25
          ||| |||||||||| ||| |||
Db    277 TCATGTGCAGGTCATCACATTG 298


RESULT  39
BG884833
LOCUS       BG884833      392 bp    mRNA            EST       30-MAY-2001
DEFINITION  dac31c11.x1 NICHD XGC Hel Xenopus laevis cDNA clone IMAGE:4408700
            3', mRNA sequence.
ACCESSION   BG884833
VERSION     BG884833.1  GI:14261925
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 392)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: dac31c11.y1
            Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Life Technologies. DNA Sequencing by:
            Washington University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            High quality sequence stop: 244.
FEATURES             Location/Qualifiers
     source          1. .392
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:4408700"
                     /clone_lib="NICHD XGC Hel"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: heart; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies."
BASE COUNT      134 a     74 c     80 g    104 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 11;  Length 392;
  Best Local Similarity   86.4%;  Pred. No. 3.7e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcagcacgtt 24
          |||||||  || ||||||||||
Db    309 GTCAGGTAGAGTTCAGCACGTT 330


RESULT  40
BG466710
LOCUS       BG466710      408 bp    mRNA            EST       20-MAR-2001
DEFINITION  EtESTed37c03.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5', mRNA sequence.
ACCESSION   BG466710
VERSION     BG466710.1  GI:13395685
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 408)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .408
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                     /dev_stage="Sporozoite stage"
                     /lab_host="SOLR"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E.tenella grown in chickens. cDNA was
                     synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to the cDNA and products were
                     size-selected on sephacryl S500. The cDNA were ligated to
                     EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
                     converted to phagemids by mass excision using ExAssist
                     helper phage and E.coli SOLR cell (Stratagene). Clones
                     were selected by negative hybridization against a pool of
                     overrepresented ESTs (N>=10,from 1682 previous reads).
                     Insert sizes range from 1.2-2.9Kb. The library may contain
                     a small percentage of host or bacterial contaminents."
BASE COUNT       99 a     77 c    114 g    118 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 11;  Length 408;
  Best Local Similarity   86.4%;  Pred. No. 3.7e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtcagcacgttg 25
          ||| |||||||||| ||| |||
Db    304 TCATGTGCAGGTCATCACATTG 325


RESULT  41
BG466711
LOCUS       BG466711      408 bp    mRNA            EST       20-MAR-2001
DEFINITION  EtESTed37c04.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5', mRNA sequence.
ACCESSION   BG466711
VERSION     BG466711.1  GI:13395686
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 408)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
```

```
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .408
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                     /dev_stage="Sporozoite stage"
                     /lab_host="SOLR"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E.tenella grown in chickens. cDNA was
                     synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to the cDNA and products were
                     size-selected on sephacryl S500. The cDNA were ligated to
                     EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
                     converted to phagemids by mass excision using ExAssist
                     helper phage and E.coli SOLR cell (Stratagene). Clones
                     were selected by negative hybridization against a pool of
                     overrepresented ESTs (N>=10,from 1682 previous reads).
                     Insert sizes range from 1.2-2.9Kb. The library may contain
                     a small percentage of host or bacterial contaminents."
BASE COUNT       99 a     77 c    114 g    118 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 11;  Length 408;
  Best Local Similarity   86.4%;  Pred. No. 3.7e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtcagcacgttg 25
          ||| |||||||||| ||| |||
Db    304 TCATGTGCAGGTCATCACATTG 325


RESULT  42
AI759462/c
LOCUS       AI759462      410 bp    mRNA            EST       18-JAN-2000
DEFINITION  EtESTea22a07.x1 Eimeria S5-2 Sporozoite stage Eimeria tenella cDNA
            3', mRNA sequence.
ACCESSION   AI759462
VERSION     AI759462.1  GI:5175213
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 410)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: ea22a07.y1
            Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 376.
FEATURES             Location/Qualifiers
     source          1. .410
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /clone_lib="Eimeria S5-2 Sporozoite stage"
                     /dev_stage="Sporozoite"
                     /lab_host="SOLR E. coli"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E. tenella grown in chickens.  cDNA
                     was synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to  the cDNA and products were
                     size-selected on Sephacryl S500.  cDNAs were digested with
                     EcoRI/XhoI and cloned into lambda Zap II (Stratagene).
                     Clones were converted to phagemids by mass excision  using
                     ExAssist helper phage and SOLR cells (Stratagene ).
                     Insert  sizes range from 1.2-2.9 kb."
BASE COUNT      120 a    110 c     71 g    109 t
ORIGIN


  Query Match             68.8%;  Score 17.2;  DB 10;  Length 410;
  Best Local Similarity   86.4%;  Pred. No. 3.7e+03;
  Matches   19;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtcagcacgttg 25
          ||| |||||||||| ||| |||
Db    129 TCATGTGCAGGTCATCACATTG 108


RESULT  43
BG413233/c
LOCUS       BG413233      411 bp    mRNA            EST       13-MAR-2001
DEFINITION  EtESTed18h02.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5', mRNA sequence.
ACCESSION   BG413233
VERSION     BG413233.1  GI:13318786
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 376.
FEATURES             Location/Qualifiers
     source          1. .411
```

/organism="Eimeria tenella"
/strain="LS18"
/db_xref="taxon:5802"
/clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
/dev_stage="Sporozoite stage"
/lab_host="SOLR"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI ; Sporozoites were obtained from in vitro sporulated and excysted oocysts of E.tenella grown in chickens. cDNA was synthesized from poly mRNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on sephacryl S500. The cDNA were ligated to EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were converted to phagemids by mass excision using ExAssist helper phage and E.coli SOLR cell (Stratagene). Clones were selected by negative hybridization against a pool of overrepresented ESTs (N>=10,from 1682 previous reads). Insert sizes range from 1.2-2.9Kb. The library may contain a small percentage of host or bacterial contaminents."
BASE COUNT 115 a 109 c 72 g 115 t
ORIGIN

Query Match 68.8%; Score 17.2; DB 11; Length 411;
Best Local Similarity 86.4%; Pred. No. 3.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 tcaggtgcaggtcagcacgttg 25
||| |||||||||| ||| |||
Db 163 TCATGTGCAGGTCATCACATTG 142

RESULT 44
BG466993/c
LOCUS BG466993 414 bp mRNA EST 20-MAR-2001
DEFINITION EtESTed31e05.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria tenella cDNA 5', mRNA sequence.
ACCESSION BG466993
VERSION BG466993.1 GI:13395968
KEYWORDS EST.
SOURCE Eimeria tenella.
ORGANISM Eimeria tenella
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae; Eimeria.
REFERENCE 1 (bases 1 to 414)
AUTHORS Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T., Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
TITLE WashU-Merck Eimeria tenella project
JOURNAL Unpublished (1999)
COMMENT Contact: David Sibley, Ph.D.
WashU-Merck Eimeria tenella project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Contact David Sibley (toxoest@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 398.
FEATURES Location/Qualifiers
source 1. .414
/organism="Eimeria tenella"
/strain="LS18"
/db_xref="taxon:5802"
/clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
/dev_stage="Sporozoite stage"
/lab_host="SOLR"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI ; Sporozoites were obtained from in vitro sporulated and excysted oocysts of E.tenella grown in chickens. cDNA was synthesized from poly mRNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on sephacryl S500. The cDNA were ligated to EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were converted to phagemids by mass excision using ExAssist helper phage and E.coli SOLR cell (Stratagene). Clones were selected by negative hybridization against a pool of overrepresented ESTs (N>=10,from 1682 previous reads). Insert sizes range from 1.2-2.9Kb. The library may contain a small percentage of host or bacterial contaminents."
BASE COUNT 119 a 111 c 74 g 110 t
ORIGIN

Query Match 68.8%; Score 17.2; DB 11; Length 414;
Best Local Similarity 86.4%; Pred. No. 3.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 tcaggtgcaggtcagcacgttg 25
||| |||||||||| ||| |||
Db 140 TCATGTGCAGGTCATCACATTG 119

RESULT 45
AI755467/c
LOCUS AI755467 418 bp mRNA EST 18-JAN-2000
DEFINITION EtESTea37c06.y1 Eimeria S5-2 Sporozoite stage Eimeria tenella cDNA 5', mRNA sequence.
ACCESSION AI755467
VERSION AI755467.1 GI:5149190
KEYWORDS EST.
SOURCE Eimeria tenella.
ORGANISM Eimeria tenella
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae; Eimeria.
REFERENCE 1 (bases 1 to 418)
AUTHORS Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T., Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
TITLE WashU-Merck Eimeria tenella project
JOURNAL Unpublished (1999)
COMMENT Contact: David Sibley, Ph.D.
WashU-Merck Eimeria tenella project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Contact David Sibley (toxoest@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 411.
FEATURES Location/Qualifiers
source 1. .418
/organism="Eimeria tenella"
/strain="LS18"
/db_xref="taxon:5802"
/clone_lib="Eimeria S5-2 Sporozoite stage"
/dev_stage="Sporozoite"
/lab_host="SOLR E. coli"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI ; Sporozoites were obtained from in vitro sporulated and excysted oocysts of E. tenella grown in chickens. cDNA was synthesized from poly mRNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on Sephacryl S500. cDNAs were digested with EcoRI/XhoI and cloned into lambda Zap II (Stratagene).

Clones were converted to phagemids by mass excision using
ExAssist helper phage and SOLR cells (Stratagene).
Insert sizes range from 1.2-2.9 kb."
BASE COUNT 119 a 115 c 74 g 110 t
ORIGIN

Query Match 68.8%; Score 17.2; DB 10; Length 418;
Best Local Similarity 86.4%; Pred. No. 3.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       4 tcaggtgcaggtcagcacgttg 25
           ||| |||||||||| ||| |||
Db     139 TCATGTGCAGGTCATCACATTG 118
```

Search completed: December 26, 2001, 12:12:29
Job time: 6511 sec